UNIVERSITY OF MAINE
LIBRARY

Class no. 330.1
Book no. Y84
Accession no. 86,416

LIBRARY
UNIV OF MAINE

LIBRARY
UNIVERSITY

ECONOMIC PROBLEMS NEW AND OLD

BY

ALLYN ABBOTT YOUNG

Professor of Economics in Harvard University

HOUGHTON MIFFLIN COMPANY

BOSTON · NEW YORK · CHICAGO · DALLAS · SAN FRANCISCO

The Riverside Press Cambridge

1927

COPYRIGHT, 1927
BY ALVIN ABBOTT YOUNG
ALL RIGHTS RESERVED

The Riverside Press
CAMBRIDGE · MASSACHUSETTS
PRINTED IN THE U.S.A.

PREFATORY NOTE

My only excuse for putting these scattered papers together into a volume is that former students and other friends have asked for such a collection. I have made my own selections, but I have tried to give weight to the opinions of others, as indicated, for example, by requests for offprints and by the extent to which other economists have thought it worth while to take account of things that I have said. The memorandum on the economic situation of Hungary in 1925 is printed for the first time, through the courtesy of Jeremiah Smith, Jr., Esq., former Commissioner General of the League of Nations for Hungary. For permission to reprint the other papers which make up the book I am indebted to the editors and publishers of the journals in which they first appeared.

Allyn A. Young

Cambridge, Mass.

86,416

CONTENTS

I. Economics and War 1

II. War Debts, External and Internal . . 21

III. Hungary in 1925 37

IV. The Trend of Prices 63

V. The Structure and Policies of the Federal Reserve System 77

VI. The Concentration of Wealth and its Meaning 95

VII. Personal and Impersonal Taxation . . . 108

VIII. Depreciation and Rate Control 119

IX. The Sherman Act and the New Anti-Trust Legislation 152

X. Some Limitations of the Value Concept . . 198

XI. Jevons's "Theory of Political Economy" . 213

XII. The Trend of Economics, as Seen by Some American Economists 232

XIII. The Measurement of Changes of the General Price Level 261

XIV. Fisher's "The Making of Index Numbers" . 276

ECONOMIC PROBLEMS
NEW AND OLD

I
ECONOMICS AND WAR[1]

"PEACE is the natural effect of trade," said Montesquieu.[2] Not is, but should be, said Adam Smith. "Commerce, which ought naturally to be among nations as among individuals, a bond of union and friendship, has become the most fertile source of discord and animosity."[3] More than a century later one of the wisest of historians reaffirmed the Scotch economist's verdict. "It is not true that the development of material interests promotes peace. Commerce, as the messenger of peace, is a mythological character. In its origin it was brigandage; in ancient, mediæval, and modern times it occasioned wars. Men fought on the Baltic for herring, and on all the seas for spices. In our day the growth of industry creates the question of foreign markets, which, in turn, brings the interests of the states into conflict. Commercial rivalry and rancor thus strengthen national hatred."[4]

In what measure is this a true finding? In what

[1] Presidential address at the thirty-eighth annual meeting of the American Economic Association, New York, December 29, 1925. Reprinted from *The American Economic Review*, vol. xvi, no. 7 (March, 1926).]

[2] *Esprit des Lois*, Book XX, chap. 2.

[3] *Wealth of Nations*, Book IV, chap. 3.

[4] E. Lavisse, *General View of the Political History of Europe* (transl. C. Gross), p. 163.

measure must it remain true? These are the questions I propose to discuss. They are difficult questions, and the literature of economics throws surprisingly little light upon them. There appear to be two general classes or types of opinions, and little else beside.

One view — which might be called the popular or naïve view — sees in war one of the normal, or even one of the rational, economic activities of men. Nations are pictured as behaving like the economic man of our methodological mythologies, each consistently and relentlessly seeking its own interests. The interests of different nations clash, conflicts arise, and conflicts grow into war.

In one way or another this view is persistently thrust before us. We encounter it in our newspapers, where it serves as one of the useful stock scenarios into which the ordinary humdrum incidents of economic and political intercourse among nations are fitted, and which give to such incidents a meaning, an element of dramatic interest, an easily recognizable place in the general course of events. We meet it in some of our books on international politics and in some of our histories — particularly in those that have been written during the last thirty or forty years. We find it in the reports by which a country's representatives in other lands keep their own government informed of matters which may be presumed to affect its interests. It even has a special literature of its own — a literature which had a mushroom growth during the war.

Such interpretations of the economic relations of nations utilize a set of familiar phrases: commercial warfare, the struggle for markets, the control of raw materials, surplus products, surplus population, economic

imperialism, economic penetration, and the like. Phrases such as the struggle for (national) existence and the survival of the fittest are also pressed into service, carrying with them the suggestion that international economic rivalries have an appointed place in the processes of nature.

There are two different ways of handling this apparatus of ideas. The characteristic of one of these methods is that nations are depicted as alert and intelligent personalities, conscious of their purposes and deliberately choosing the means of achieving those purposes. Their foreign policies are conceived in terms of strategy, of calculated economic advantage. They are, in short, rational, or Machiavellian, states.

The other way of dealing with this common stock of notions uses the apparatus of historical determinism. Causes, not purposes, rule. National policies and national destinies are shaped by the cumulative pressure of antecedent forces. The fruits of this method are a variety of pseudo-scientific dogmas respecting the "ultimate" causes of war. I call them pseudo-scientific because, although they borrow the mask of science, they are and must be arbitrary and unverifiable.

Of these various dogmas the one which has had the most pervasive influence is the Marxian. This doctrine is to the effect that changes in industrial technique, coupled with the private ownership of those instruments of production through which alone the fruits of technical progress can be realized, bring about a disparity between a country's power to produce and its power to consume. The surplus product accumulates in the market, where it leads periodically to crises. Endeavors to find an outlet for it lead to economic rivalries among nations, to the exploiting of undeveloped or back-

ward countries, to colonial expansion, to economic imperialism, and to wars.

The rational state does not appear in this picture. Instead there is the capitalistic or class state, itself only an historical incident, marking a certain stage in the development of economic and political institutions. The abiding thing is the conflict of the interests of different economic classes. Modern wars are undertaken to secure the advantage, not of a nation, but of a class. The real lines of cleavage are to be found, not at national frontiers, but in the horizontal stratification of society.

One element in this Marxian doctrine, namely the notion of surplus products pressing outward across national boundaries under military convoy, has an especially wide vogue. Writers who do not count themselves followers of Marx, as well as those who do, employ it as a routine formula.

The second general class or type of opinions to which I have referred is distinguished, not by a special emphasis upon some particular view of the nature and purposes of the aggressive activities of states, but by a very definite thesis with respect to the wisdom and the consequences of such activities. If wars are waged for economic advantage, it is held, they defeat their own purposes. So, too, in general, with all national policies designed to advance the economic interests of one state at the expense of other states. The truth is, it is alleged, that a nation gains by the prosperity of other nations, not by their poverty.

This general thesis, if stated with some necessary qualifications, would be subscribed to by most economists. It was brilliantly expounded in Mr. Norman An-

gell's book, *The Great Illusion*. If in the days of its first vogue that book seemed to be given little attention by the economists, it was not because they disagreed with its conclusions, but rather because most of those conclusions seemed to them to be fairly commonplace economic doctrines. Doubtless Mr. Angell weakened a good case by pushing it a little too far. He gave too little weight to the special interests (not necessarily or even generally class interests) that may be served by a belligerent or imperialistic policy, even when other interests, larger but more diffused, are injured. He did not adequately distinguish between immediate and ultimate gains and losses. But taking his argument in the large, it would command, I believe, the general assent of economists. Some of the policies he finds unwise are, in fact, policies economists are accustomed to disparage by lumping them together and calling them neo-mercantilism.

Assuming that Mr. Angell and the economists are right, is there ground for hope that, as the result of a slow process of education, the world would become convinced that aggressive economic nationalism is profitless? Is there reason to believe that in this event the so-called economic causes of war would be done away with? Such hopes, we may be fairly sure, would be vain. For one thing, we have learned in other fields that progress which has to wait upon men's becoming more reasonable is likely to be delayed indefinitely. For another thing — and here I pass to one of the central themes of this paper — there is ground for challenging the common views of the nature of international economic competition and of its relation to war. The facts are too complex and too tangled to be fitted into the concepts and formulas which we ordinarily use.

We get those concepts and formulas, for the most part, from the market place. We use them in describing and analyzing the mechanism of money-making, of business competition. Here we find uniformities, rational rules of conduct, out of which we build the framework of economic science. That reliable mechanism, the "economic man," is merely the average man, taken in his business relations — taken, that is, as buyer and seller. Now, save under very exceptional conditions the economic relations of national groups are not like those of buyer and seller. Not since the days of the crudest types of mercantilism have they been so regarded by economists.

We concern ourselves with what nations *should* do in order to secure their maximum economic advantage, we take account of some of the things they actually do, and we even venture to explain or find reasons for certain of these activities. But unless we are the bigoted devotees of some dogmatic philosophy of history, we do not expect that the behavior of national groups will conform to some rational rule, that it will be stable and predictable.

The values of the world of international rivalry are more like the irrational values of the world of consumers' choices than they are like the money profits and the other money incomes for which men contend in the world of commerce. Consider such phrases as economic dominance, empire, economic independence, a place in the sun, territorial expansion, control of markets, freedom of the seas. These phrases denote some of the things for which men are supposed to fight. Each has an economic significance. And yet, what discernible relation is there between their potency for war and their economic significance? Their real meaning appears

only when they are projected against an historical background; but there they lose any peculiarly economic quality, and become merged in the general picture of national prejudices and passions.

What I am trying to say has been put more skillfully by Walter Lippman: "How does it happen that the people not concerned in a special interest are so ready to defend it against the world? . . . The most obvious reason is that the private citizens are in the main abysmally ignorant of what the real stakes of diplomacy are. They do not think in terms of railroad concessions, mines, banking, and trade. . . . Each contest for economic privileges appears to the public as a kind of sporting event with loaded weapons. The people wish their team, that is, their country, to win. . . . Business is the chief form which competition between nations can assume. To be worsted in that competition means more than to lose money; it means a loss of social importance as well. . . . The way to increase national prestige is to win economic victories by diplomatic methods. . . . Armament is added as an 'insurance' for diplomacy, and of course military preparation always calls forth military preparation. Every international incident is seen then, not on its 'merits,' but in its relation to the whole vast complicated game, forever teetering on the edge of war." [1]

It may be that some of the interests which arouse these belligerent emotions and around which they cluster are the rational economic interests of different national groups. It is plain that some of them are the real interests of particular men or particular classes. But, as I have said, there is no correlation between the economic importance of these interests and their power to rally a

[1] *The Stakes of Diplomacy*, pp. 76–83.

people to their defense. They may be wholly factitious, and yet be potent.

There is instruction to be had from recent developments in the field of sports. I mean, of course, intercollegiate and other inter-group sports, in which small teams of selected competitors carry with them into their contests the rivalries of the groups they represent. Members of these rival groups like to absorb themselves in these conflicts by reading about them. The purveyors of sporting news have found that this is an appetite which can be stimulated, that the interests of persons outside the rival groups may be engaged, and even that the number of partisans may be increased.

They have found also that the number of *competitions* may be fruitfully multiplied by bringing existing contests into new relations. An intricate system of mythical regional and class "championships" has been invented. Teams are ranked by the percentage of victories, by the total scores they have made, and in various other ways. Ingenious methods of rating the achievements of individual players have been devised.

These artifices succeed in entrapping the interests of readers. And they accomplish more than that. Some of these make-believe competitions become real. The spirit of rivalry reaches out and takes hold of them. They supply new criteria of superiority, new symbols of prestige. Some of them come to be regarded as expressions of inevitable "natural rivalries." The facts, however, suggest that though there may be a vaguely circumscribed field of potential rivalries, within that field competitive alignments are free to arrange themselves in various patterns, while one thing or another may come to be adopted as an emblem of success.

In the larger field of international economic rivalries group psychology retains its characteristics. The world in which national groups strive to realize their opposed interests is in large part a world of man-made patterns and symbols. It is a new world, for nations as we know them are only a few hundred years old — no older than the new world-commerce which helped to bring them into being and which remains one of the fields in which their oppositions and rivalries feed and grow.

Among the architects of this world have been the historians. They have played a rôle not unlike that of the contemporary newspaper annalist of competitive sports. Most histories have been histories of nations. By abstracting the nation from the other forms which human relations take, by emphasizing the peculiar and differentiating elements in a nation's institutional heritage, they deepen the cleavages between national groups. Many histories have put disproportionate emphasis upon wars. And in explaining wars they have often put more reliance upon "economic motives" than either economics or psychology would warrant.

"It is possible to study a multitude of histories," says John Dewey, "and yet permit history, the record of the transitions and transformations of human activities, to escape us. Taking history in separate doses of this country and that, . . . we miss the fact of history and also its lesson; the diversity of institutional forms and customs which the same human nature may produce and employ. An infantile logic, now happily expelled from physical science, taught that opium put men to sleep because of its dormitive potency. We follow the same logic in social matters when we believe that a war exists because of bellicose instincts; or that a particular economic regime is necessary because of acquisitive and

competitive impulses which must find expression. . . . We have constructed an elaborate political zoölogy as mythological and not nearly as poetic as that other zoölogy of phœnixes, griffins, and unicorns. Native racial spirit, the spirit of the people, or of the time, national destiny are familiar figures in this social zoo. As names for effects, for existing customs, they are sometimes useful. As names for explanatory forces they work havoc with intelligence." [1] And so, we may add, do such names as economic motives and the economic causes of war.

William James wrote a famous essay on the "The Moral Equivalents of War." Conceivably we might speak of equivalents or substitutes for economic antagonisms. But what we need most are substitutes for the habitual patterns which elicit and direct those antagonisms.

Something may be accomplished, undoubtedly, by continued emphasis upon the wasteful stupidity of most of the efforts national governments make to secure economic advantages outside of their own territories, although this something is not very much. There is reason to hope, moreover, that with the increase in the number and variety of contacts between the peoples of different countries, national antagonisms will diminish. Not, however, that the growth of concreter forms of knowledge, born of such contacts, can be counted upon to dissipate our abstract notions of alien types of men and of rational but malevolent states. The real ground for hope is rather that with the growth of communication and of economic interpenetration, new forms of organization will have to be devised

[1] *Human Nature and Conduct*, pp. 110–12.

Organization generally provides a mechanism by means of which some measure of external control can be exercised over men's activities. But that is only a small part of its significance. Organization *directs* activities as well as controls them.

Political organization has not kept pace with economic organization. Increasing interdependence asserts itself in economic life. Raw materials, markets, borrowing and lending, trade routes, prices, monetary and banking policies are things in which the different peoples of the world have a joint as well as a separate interest.

National states, each acting only for itself, are inefficient guardians of these joint interests. Within a nation's own boundaries it manages to bring the conflicting interests of different sections and groups into some sort of balance and to enforce general standards and rules governing the conduct of business enterprise. Just because there are these rules of the game, because political organization is nation-wide, the game takes on a different character, the nature and the meaning of sectional conflict is changed. An American scholar who has thrown a new and transforming light on our national history has said: "We must frankly face the fact that in this vast and heterogeneous nation, this sister of all Europe, regional geography is a fundamental fact; that the American peace has been achieved by restraining sectional selfishness and assertiveness and by coming to agreements rather than to reciprocal denunciations or to blows. . . . Statesmanship in this nation consists not only in representing the special interests of the leader's own section, but in finding a formula that will bring the different regions together in a common policy." [1]

[1] F. J. Turner, "The Significance of the Section in American History," *Wisconsin Magazine of History*, vol. XIII, pp. 275, 279 (March, 1925).

But the nation, in its larger relations, is itself a section. Outside of its own borders its interests, real or supposed, conflict with the interests of other nations. Here statesmanship becomes sectional leadership. The rôle which the statesman finds easiest to play is often one which is assigned to him in the popular dramatization of the facts of international economic intercourse. His task would be harder, however, if his countrymen insisted or expected that the foreign policy of their own government should have the same degree of design and purpose, the same elements of plot and strategy, that they see in the foreign policies of other states.[1]

There is a modern animism which imputes malign intent, not to the forces of nature, but to personified "nations." The preposterous myth that Pan-Germanism was a definite national policy could not have found lodgment except in minds patterned to receive it. In a newspaper which lies before me as I write, there is a short dispatch telling that a new German company has been organized which proposes to operate in Russia. The headline reads: "Germany Getting Grip on Russia." Not long ago another dispatch told of a rather notable increase of some relatively unimportant German exports to Sweden. The headline and an introductory paragraph interpreted this ordinary incident of trade as meaning that Germany was getting "control" of Swedish markets. And when it is not Germany it is England or Japan or some other country. A clipping bureau, I

[1 "Whoever has been inside British foreign policy is familiar with the emotion of indignation, amusement, or contempt with which he reads of the deep motives and the clever schemes that are invented for present-day British diplomatists and attributed to them by ingenious writers in foreign, and sometimes even in the British, press. One who is conscious of this may well be cautious in attributing deep and sinister designs to the action of foreign Governments." — Viscount Grey, *Twenty-Five Years*, vol. I, p. 230.]

suppose, could supply hundreds of such items each year. No one country has a monopoly of these childish absurdities. Running through the pre-war files of an important German commercial journal I found an astonishing number of references to the "conquest" (*Eroberung*) of Latin America by the United States; and that interpretation of our plans (not of our achievements) is common in Latin American countries.

A good deal of what passes for information respecting the purposes and activities of other nations is no better than malicious gossip. Such is the character of some of the information which governments receive through diplomatic channels. Bismarck said, "I have often not shown dispatches from our representatives in German Courts in the highest quarters, because they had a tendency to be piquant, or to relate and give importance to annoying expressions or occurrences, rather than to foster and improve the relations between the two courts, so long as the latter, as in Germany is always the case, was the task of our policy." [1] But even when a government's policy is "to foster and improve relations," no such censorship is possible over the open channels of communication through which the people of a democracy gain their impressions of the purposes of other states.

However slight their foundation, these impressions, like our general views of the nature of international economic competition, help to determine our attitudes and to shape our conduct. Fear of the power and the purposes of American trusts was one of the reasons Germans gave for the organization of cartels in some of their export industries. The German cartels, in turn, figured largely in the discussions which led to the Webb-Pome-

[1] *Reflections and Reminiscences* (Eng. transl.), vol. II, p. 248.

rene Act of 1918 — a particularly invidious piece of legislation, which permits in American export trade combinations of a type that is illegal in domestic trade, with the proviso that such combinations shall not "restrain" the trade of or compete unfairly with *American* competitors.

In a hundred other incidents where retaliation leads to retaliation, any one who cares to look may see part of the actual process by which a world of gossip and of myths becomes the world of national policies. Nothing else is possible so long as each separate state is not only the guardian of its own interests but is also the interpreter of the interests and purposes of other states. There is no more ground for hoping that these types of economic friction will be done away with by changes in the policies of separate states than there is for expecting that armaments can be reduced effectively in any other way than by common agreement among nations.

In certain limited fields of economic activity, common agreements have already been reached and organs of international administration have been established. This is notably true in respect of communications and transport, where elementary considerations of convenience and economy, as well as the common interest in uniform and nondiscriminatory rules, make agreements imperative. The advantages of conformity are plainly visible, while there is little in nonconformity to which even a fictitious "national interest" can be attached.

There is more significance in what has been accomplished in freeing transit trade from duties and from unreasonable transport charges and in creating international easements in important rivers and canals.

Here the interest of a particular state may sometimes be opposed to the interests of other states. Taking advantage of a favorable geographical position, it might take tribute from the commerce which crosses its territory (as states often have in the past). There has been an element of compulsion — sometimes the compulsion of stronger states and sometimes the compulsion of circumstances — in the progress that has been made in these matters. To take only one example, the provisions relating to freedom of transit in the treaties concluded at the end of the World War, one-sided though those provisions were, helped to prepare the way for the general European agreement embodied in the Barcelona Convention of March, 1921.

Agreements for the suppression of the use of unfair methods of competition in international trade have been limited, for the most part, to the protection of patents, copyrights, trade-marks, trade-names, and the like. In these matters the interests of different states are not so much joint as they are reciprocal, and they are unevenly reciprocal. Substantial inequalities remain, particularly in the effectiveness with which the provisions agreed to are enforced.

International agreements with respect to labor legislation may possibly be regarded as tending toward fairer standards of international industrial competition. But that is not their chief significance; and I can do no more than mention them here. Nor can I discuss the important preparatory work done under the auspices of the Economic Committee of the League of Nations on commercial arbitration, export and import prohibitions, double taxation, and other matters.

It will be observed that few of the operative agreements which I have discussed thus far cut deeply into

what are generally deemed to be important national interests or have come to be symbols of national prestige. None of them has much news value. The history of the international agreements that have sought really to limit the scope of international commercial "strategy" has been a record of halting achievement. Like price agreements and pooling arrangements among business concerns, such covenants appear to remain effective only so long as that situation serves the interests of the more important signatories. Such was the history of the Berlin Act of 1885, of the Brussels Sugar Convention of 1902, and of the Act of Algeciras of 1906.

In the absence of general agreements, the international economic relations of a large part of the world have been governed by a complex network of bilateral commercial treaties, held together and made at all consistent and tolerable only by most-favored-nation clauses. The system thus set up, especially in continental Europe, was unstable; for a single important new treaty would upset things until, by the revision of other treaties, a condition of temporary equilibrium could again be reached. The system was uneven; for the strong states were able to bargain more effectively than the weaker ones. In particular, industrial states had a more advantageous position than agricultural states. Questions of commercial policy were recurrently projected into the field of international politics, so that they were quite commonly discussed in terms of *Machtpolitik*. Commercial treaties were regarded as instruments by which a state's power could be projected across its own boundaries.

Such a system often has undesirable economic effects. The present tariffs of some of the new states of eastern Europe afford an illustration. These tariffs are much

higher than the real economic interests of those states demand. Artificial barriers set up at new boundary lines, these new tariffs block long-established channels of trade and hinder the economic recovery of that part of the world. Not all the blame can be put upon the newly stimulated spirit of nationalism. For these are bargaining tariffs. Some of the highest duties are imposed, not upon goods which the tariff-making country particularly desires to keep out, but upon goods which another country desires to get in. Aimed at one another, these new tariffs are an example of wasteful futility. Commercial treaties, coupled with most-favored-nation provisions, will gradually reduce them. In the meanwhile the costs are heavy. And experience has shown, I think, that tariffs made in this way generally have an upward trend. New vantage points are sought from time to time; special industrial interests look upon the reductions made in treaties as ground regrettably lost; national sentiment, also, learns to look upon concessions as retreat, so that there is a cumulative pressure upward.

The United States has stood aloof from this system. We have held, though not with complete consistency, to the policy of the equal treatment of the commerce of all other nations. But the Tariff Act of 1922 empowers the President to impose retaliatory duties upon imports from any country which discriminates against our commerce; and the same general principle was embodied in the Tariff Act of 1909. We determine the other country's guilt; and we impose what we consider an appropriate penalty.

Students of these problems have become pretty well convinced, I think, that the most important single step toward their solution lies in the adoption of multilateral

conventions defining the conditions of economic intercourse among the nations of the world. The first principle, the irreducible minimum, in such agreements is the doing away with a nation's power to discriminate against the trade of any other nation. This was the meaning of the third of President Wilson's Fourteen Points: "The removal, so far as possible, of all economic barriers and the establishment of an equality of trade conditions." This, beyond doubt, is part of the meaning of the clause in the Covenant of the League of Nations which binds the signers of that covenant to "make provision to secure and maintain freedom of communications and of transit and equitable treatment for the commerce of all Members of the League." [1]

The League's word "equitable" is better in some ways than President Wilson's word "equality." It does not carry with it so clear and definite a commitment; but for that very reason it does not suggest precise limits to the scope of international economic agreements.

In fact, the phrase, "equitable treatment for commerce," suggests different things to men in different countries. To some it means equal, that is, general most-favored-nation treatment. Some would add national treatment in respect of certain matters, such, for example, as the rights of foreign vessels in national ports, the taxes imposed upon foreigners, or the protection given them against unfair competition. To others it includes anti-dumping arrangements, the removal of restrictions on exports, unrestricted access to raw materials, or the open door in undeveloped parts of the world. But just now the content of international

[1] I have given an account of the history of this clause in *A History of the Peace Conference of Paris*, edited by H. W. V. Temperley, vol. v, chap. 1, part 3.

economic agreements is not so important as it is that agreements should be reached.

I have already said that the more important agreements of this general sort made before the war were unsuccessful. The difficulty with them was that they were *exceptional* arrangements. Belgium in the Congo and France in Morocco were hampered by restrictions that had the character of special disabilities, such as had not been attached to the colonial expansion of other countries. The Brussels Sugar Convention was an anomaly in a world in which nations were generally left free to determine their commercial policies as they pleased. To be really effective, such agreements must establish a general rule, not an exception to a rule. They must in some way create a new standard pattern of thought and conduct.

The resolution by which a few months ago the Assembly of the League of Nations invited the Council of the League to institute preparations for an international economic conference referred to the "economic difficulties which stand in the way of the restoration of general prosperity," and expressed the conviction that "economic peace will contribute largely to ensuring the security of peoples." As matters now stand in Europe, these two objects, economic restoration and economic peace, are closely connected. But of the two, economic peace is the more important, for it is the condition of the other.

Consider what has already been accomplished toward financial reconstruction. The real significance of the plans for the financial rehabilitation of Austria, Hungary, and Germany is that they embody international agreements. The securing of these international agreements was both a more difficult and a more important achieve-

ment than the formulating of the particular economic remedies that were to be administered. It is easier to determine what economic procedure is wise than it is to change national attitudes.

In these instances the changing of national attitudes was achieved by organizing and giving expression to the common interests of nations. The hopelessness of other methods had to be shown before this could be accomplished, but once done, it is not easily undone. The new attitude creates a new interest. Even if some of its economic provisions should break down, as is altogether likely, the Dawes plan probably will have solved the reparation problem.

The significance of these achievements for the general problem which I have undertaken to discuss is obvious. The attitudes and activities which we have in mind when we speak of "the economic causes of war" are not inevitable and unyielding expressions of permanent traits of human nature. They are forms or patterns of conduct and are correlated with particular modes of organization. Other forms and patterns, associated with other modes of organization, are within the bounds of practicable achievement. This does not mean that the task is simple, or that it can be accomplished merely by finding a magic formula. No sensible person expects that sectional interests or international economic antagonisms will disappear. But it is not unreasonable to hope that some day they may be subordinated to new and larger interests which will grow out of new forms of organization. At any rate, the only way to secure economic peace is to turn our eyes towards it.

II

WAR DEBTS, EXTERNAL AND INTERNAL[1]

How far is it true that a large domestic or internal national debt, as contrasted with an external or foreign debt, is a negligible burden? Does the payment of an internal debt, requiring merely the transfer of wealth or income from some persons to others *within* the same country, put no strain upon that country's economic energies? How far is it true, by way of contrast, that Germany's capacity to make reparations payments, or the capacity of any nation to make large foreign payments, is rigidly limited by its ability to maintain a favorable balance of commodity exports over commodity imports?

At first the very real differences between the burdens imposed by external and by internal debts were ignored in popular discussions and flouted in the policies of governments. That fact explains and justifies the emphasis competent critics, in increasing number, have put upon those differences. But despite the great value of the educational work they have done, some of these critics have drawn too straight a line between the two kinds of debts.

We may find a helpful approach to the problem if we look first at several different ways of appraising a country's ability to pay a heavy and burdensome foreign debt.

In the first place, there is the now familiar method which relies upon an inventory of a country's present

[1] Reprinted from *Foreign Affairs*, vol. II, no. 3 (March, 1924).]

foreign assets coupled with an estimate of its capacity to maintain a favorable balance of trade through a series of years. This latter estimate is generally based upon an analysis of the country's pre-war trade and its possible rate of growth, together with a consideration of the degree in which the war has affected the country's productive resources and its markets. This is the method used by Mr. Keynes in *The Economic Consequences of the Peace*, and by Messrs. Moulton and McGuire in *Germany's Capacity to Pay*. In competent hands it leads to important results. Despite the highly conjectural character of some of the figures that must be drawn upon, the better estimates that have been made in this way are not far apart.

Such studies of a debtor country's possible balance of trade are indispensable. They are a necessary preliminary to any reasoned judgment upon that country's capacity to make foreign payments. But it is wrong to regard them as complete and adequate estimates of capacity to pay. They make no room for the play of various elastic factors. In particular they take no account of the readjustments in the economic and financial relations of the different nations of the world that are bound to be brought about as a result of the very operation of paying a heavy foreign debt.

In the second place, we may turn to the arguments of those who hold that the effort to pay a large foreign debt has the curious effect of greatly increasing the debtor country's real capacity to pay — provided that the currency of the debtor country is inconvertible paper.[1] Under such conditions, it is even contended, a country's

[1 What follows has little bearing upon the problem of the effect an effort to pay a large foreign debt has upon the export and import trade of a country which maintains a gold standard.]

ability to make foreign payments is really limited only by its ability to produce more than it consumes. The gist of the argument is as follows: To make payments on its external debt the Government must buy foreign bills of exchange in large quantities. The price of such bills, and along with it the prices obtained (in domestic currency) for exported goods, will be pushed up rapidly, so that finally a considerable differential will be established between the general domestic price level and the prices that can be had for exports. This differential operates, so the doctrine runs, virtually as a bonus on exports. It will induce business men to turn a larger share of their energies to producing goods for the export rather than for the domestic market. Why should German producers, for example, sell goods at home when more marks could be got for them by selling them (or other goods produced at no larger cost) abroad for credits in dollars or pounds or francs which could then be sold to the Government at high prices? So long as this bonus on exports can be maintained by the pressure of the Government's continuing demand for the means of making foreign payments, so long, in a quasi-automatic way, the country's production of goods for the export market will be increased, while in equal measure its production of goods consumed at home will be decreased. Its capacity to pay, therefore, will be limited in the long run only by its maximum productive capacity on the one hand and its minimum domestic industrial and subsistence needs on the other.

Now this theory is based upon a perfectly sound principle, but it runs that principle into the ground by fantastically exaggerating its possibilities. The first of the two methods we have considered, taken by itself, is too inelastic. This second method errs in the other direc-

tion. It ignores a number of very important inelastic factors. It takes no account, for example, of the inelasticity of the demand of the world's markets for the exports of any one country or of the inelasticity of that country's demand for imports. It passes over the fact that a country like Germany must shape its exports very largely from imported raw materials and that the prices that must be paid for imported goods, in domestic currency, generally rise at least as rapidly as the prices that can be realized for exports. It disregards the way in which such price differentials react upon the whole structure of industry and finance within the debtor country. It forgets the inevitable lowering of the standard of living and the disintegrating effects of maladjustments in the delicate interrelations of the country's system of prices. Financial and industrial wreckage, rather than an increasing export surplus, is the certain result of the pressure of an unduly large foreign debt.

Furthermore, even though some of the advantages claimed might be secured by export industries if export and domestic prices and the differential between them could be maintained at a moderate and stable level, such advantages would disappear when, as would be inevitable, the export and domestic price levels continued to advance. An *increasing* differential between export and domestic prices will not always stimulate exports. Some exporters will prefer to wait for a higher differential later. Some importers, on the other hand, will be induced to bring in foreign goods, at whatever cost, in order to sell them later when the margin between the prices of imports and the general level of domestic prices is larger.

A third method of approach to the problem of the maximum manageable size of a foreign debt is suggested

by the proposals that Germany's payment should be made in a lump by turning over shares in German industries to Germany's creditors. There have been some such proposals, on the part of Germans, that have had an official or semi-official character. A distinguished German statistician, Dr. R. R. Kuczynski, has gone even further, proposing a heavy capital tax upon German property of whatever sort, the proceeds, in the form of mortgages, bonds, shares, and the like, to be handed over *en bloc* in acquittance of reparations obligations.

It is going too far to hold that such methods of payment "do not constitute actual payments at all." [1] I do not suppose that it is claimed by Dr. Kuczynski, or by others who have made similar proposals, that a final or *economic* payment could be achieved by such a transfer. A *legal* payment might conceivably be made and a receipt in full secured by assigning or transferring securities which would themselves be nothing more than evidences of debts. The contention merely is that the final economic settlement of the debt could be reached more easily and more efficiently if it were removed from the field of international politics, and if the mechanism of the payment were controlled by the free play of economic forces, rather than by the attempts of governments to collect the payments in one form rather than in another.

There is no reason to doubt the sincerity and good intentions which prompt these proposals, and it is in some respects a matter for regret that any large reliance upon such methods is neither wise nor practicable. The difficulties, both political and economic, are manifold. For one thing, the assets thus secured, if large

[1] Moulton and McGuire, *Germany's Capacity to Pay*, p. 18.

and miscellaneous, would shrink greatly in value in the markets of the world.

Back of such proposals, nevertheless, there is a thoroughly sound notion which supplements and corrects the rigid export-surplus doctrine. If the reparations debt were fixed at a reasonable amount and if Germany were left free to find her own ways and means of payment, it is fairly certain that a situation would develop in the end which would somewhat resemble — though on a reduced and moderate scale — the situation which these proposals contemplate as the result of a single operation. By the time the reparations payments were completed, other countries — not Germany's present creditors alone — would hold considerable amounts of German securities, both public and private. In various ways, planned or spontaneous, a fairly large proportion of the reparations debt would be refunded into these other forms. Legal payment would be completed and the German Government acquitted of its liability long before the final economic payment had been made. Indeed, the continual shiftings of the world's balances of international debts and payments might make it impossible to say with certainty at any one time that the final payments *had* been made.

It is likewise difficult to think that France will receive any large part of her share of reparations in the form of such manufactured goods as Germany might produce. It is distinctly more probable — her own external debts aside — that France's increased imports, like Germany's increased exports, would be diffused over the whole surface of world trade, and that her foreign holdings of various sorts, her investments in other parts of the world, would be measurably increased, although not necessarily in exact proportion to the increase of Ger-

many's "refunded" foreign liabilities. Such changes in the world's international balance sheets would, beyond doubt, react upon and modify the currents of international trade in commodities — but the discussion of such matters would take us far afield.

All of the differences between external and internal debts are bound up with the fundamental fact that the former call for foreign payments. Such payments create problems in foreign trade and foreign exchange. So long as the payments continue, the debtor country, except so far as it can refund the debt, must produce more than it can consume or add to its accumulated domestic capital. The monetary units in which payment must be made and which therefore measure the amount of the debt are not within the arbitrary control of the debtor government. Domestic inflation makes these foreign monetary units dearer, not cheaper. Deliberate repudiation, complete or partial, of external debts is not as a rule practicable. Such are some of the more important characteristics which external debts have but which internal debts lack.

Let us turn now to points which they have in common. We note, in the first place, that the payment of an internal, as of an external debt, requires taxation, calling for sacrifices and exercising a repressive effect upon trade and industry.

In the second place, there is an analogy — not merely fanciful, although it would be easy to push it too far — between a country's taxpayers and the holders of its internal debt on the one hand, and debtor and creditor countries on the other. Taxpayers and holders of the public debts are in part two different groups. The interest and principal of the debt enter into the balance of

payments as between the taxpayers and the owners of government securities. For many years the taxpayers, as such, must produce more than they can consume or save, while the Government's creditors will receive more than, at the time, they are producing. If the debt absorbs any large proportion of the national income, as in France, England, and other European countries, the resulting reactions upon the distribution of incomes, upon the general standard of living, upon the demand for goods, and upon the general structure of industry may be considerable. Much depends, of course, upon the way in which the burden of taxation is apportioned, upon whether taxes fall more or less heavily upon necessities, upon luxuries, upon incomes, or upon accumulated capital. But at the best the strain upon the country's economic activities is likely to be heavy, even though the complications and special difficulties attending transactions in foreign exchange are absent.

In the third place — and this I believe to be the most important point of similarity — heavy internal debts, like external debts, strike at the heart of a country's economic life by bringing disorder into its currency. That inflation, unbalanced budgets, and disordered exchanges have been among the chief factors delaying Europe's economic recovery is pretty generally known, but it is not so well understood as it should be that the sequence of cause and effect, particularly in the period following the war, has not been inflation, unbalanced budgets, disordered exchanges, but unbalanced budgets, disordered exchanges, inflation.

Before the war the world had an international money — gold — which not only stabilized (albeit imperfectly) but tied and held together the currencies of the great commercial countries. The mechanism of international

payments operated with deceptive ease and smoothness. The price system was, in very large measure, international. The market for loanable funds, whether of long or short maturity, was also, in a very considerable degree, international.

At present a large part of the western world is cut up, artificially, into separate economic districts within which purely national curriences rule. The worst thing about these new artificial barriers is that they fluctuate continuously and uncertainly — inviting hazardous speculation, but discouraging the steady flow of trade.

The war was financed, as we know, partly by taxes and to some extent by voluntary savings, but mostly by diluting the purchasing power of money. Inflation was the mechanism by which war debts, we might say, were created. All this is now very generally understood. But there appears to be no adequate appreciation of the intimate and necessary connection between debt reduction, on the one hand, and the deflation and stabilization of paper currencies, on the other hand.

This relation between debts and the status of paper currencies holds whether the debts be external or internal, although the mechanism by means of which the increase or decrease of the debt operates upon the currency is different in the two cases. Consider, for a moment, some of the monetary effects of a burdensome foreign debt. The wrecking of Germany's monetary system affords an instructive example.

The Government's demand for foreign funds was added to similar demands on the part of importers and of others having foreign payments to make. The only way of securing foreign exchange was to bid its price up to a point where some of the competing demand was

eliminated. The price of dollars, of sterling, and of francs went up out of all proportion to the difference between the domestic purchasing power of marks within Germany and the domestic purchasing power of the other currencies in the countries in which they circulated. But the prices in marks of imported and exported goods — of goods, that is, for which there is an international market — were driven rapidly up by the advance in the price of foreign exchange. Tied to the prices of the international market in a thousand indirect and intricate ways, the domestic price level, though until recently lagging considerably behind, advanced haltingly but irresistibly. Inflation, though accelerated by an unbalanced budget, followed — it did not precede — the advance of prices. During the past four years the per capita circulation of money in Germany, measured in terms of its domestic purchasing power, has been much smaller than before the war. Measured in terms of its gold value the shrinkage of the currency has been even more striking. In large measure inflation has been the result rather than the cause of the depreciation of the value of the currency.

I have reviewed these matters again in order to give point to the emphasis I shall put upon what I believe to be the indispensable key to an understanding of the vagaries of the behavior of the depreciated currencies of Europe. I refer to the dominating part played by speculation. It makes all the difference in the world whether men's actions are prompted by the belief that a depreciated currency will increase in value or by the belief that its future trend will be downward.

The turning point in Germany's financial fortunes came in the first half of 1922, following upon the adverse decision with respect to Northern Silesia, late in the

autumn of 1921. Up to that time many people in and out of Germany had maintained a persistent if unreasoned faith in the future of the mark. That circumstance measurably retarded the inevitable decline of the mark, and did much to lighten Germany's financial burden. It made exports of German goods larger than otherwise they would have been. It led people in other countries to make large speculative purchases of marks and of securities payable in marks. It was one of the things that induced persons in other countries to invest heavily in German property. During that period, it is safe to say, Germany's imports of capital, in various invisible forms, were distinctly larger than her invisible exports. Part of the favorable balance was available for reparation payments.

This misplaced confidence in the future of the mark waned rapidly after the failure of the attempt to stabilize its value in December, 1921, and a complete reversal of the general trend of speculation followed. With a bear market for marks the difficulties of the German Government were multiplied. Then came the unrelenting drift of events into the reparations crisis of a year ago, with its costly and ominous sequel.

With the conviction once established that the value of the mark, in terms of other currencies, was bound to continue to depreciate, its depreciation was accelerated. Commodity exports were retarded and imports stimulated. The direction of the current of invisible elements in the balance of international payments was reversed. Capital began to flow out of Germany more rapidly than into it. The rapid decline of the mark under the pressure of these forces only added to their strength. The mechanism of financial disintegration worked in a cumulative way.

Some absurdly exaggerated stories are afloat respecting the amount of German capital that has gone during the last year and a half into bank deposits and into the purchase of securities in the United States, Canada, Switzerland, and other countries with relatively stable currencies. But underneath these exaggerations there is reasonable ground for presuming that Germany's exports of capital have been large.

We do not have to attribute these developments to a concerted plot on the part of German business men, abetted by the German Government, to evade their reparations obligation. It is reasonably certain, of course, that the fear of drastic taxation helped to create a preference for foreign holdings. But the major compelling force back of the export of German capital has been the loss of confidence in the mark. The particular methods by which such transactions are effected do not matter greatly. The general result is that the foreign credits created by exports and in other ways are allowed to remain abroad instead of being offered for sale to the Government. In any case the transaction is essentially a bear operation in marks. As such it is easily explainable. It is wrong only in so far as it involves a violation or evasion of the laws regulating the purchase and sale of foreign bills of exchange.

I cannot believe that the real attitude of the French Government toward this matter, or, at any rate, the attitude of the economic advisers of the French Government, has been correctly reported, or that it is adequately expressed in the official statements of that Government. A very definite impression has been given that the French Government objects to the removal of "wealth" from Germany, and that it insists that ways and means of bringing it back again must be

found. This, it appears, is a problem upon which one of the committees of experts now attached to the Reparations Commission is at work. On the face of it this attitude is paradoxical, for so long as German wealth remains in Germany it can be used only with difficulty for reparations payments. German-owned dollars, francs, and sterling, on the other hand, are precisely the stuff out of which reparations payments can easily be fashioned.

The real ground of the French complaint must be that these privately owned foreign funds, which have been accumulated at such heavy cost to Germany, might, at no greater cost, have been acquired by the German Government and then applied to the payment of reparations. The French Government does not really desire that these foreign funds be brought back into Germany in the form, let us say, of imports of merchandise. Its real and legitimate interest is that the German Government should be able in some way to possess itself of these foreign assets and turn them over to Germany's creditors.

It is impossible that this legitimate end should be obtained by ferreting out these foreign holdings or by any amount of pressure and intimidation. But it would be accomplished promptly and easily if a stable monetary system could be established in Germany. With the motive for the concealment and expatriation of funds destroyed, the German Government could get hold of them by paying a fair market price in German currency.

It will be granted, perhaps, that the history of the German mark illustrates and confirms the principle that a stable currency and an excessive external debt are incompatible. A similar relation holds as between a

country's currency and an excessive internal debt, although only a part of the mechanism by which external debts affect the price level operates when the debts are internal. An excessive internal debt makes it difficult to balance the budget, and an unbalanced budget calls for further advances by the banks, or for further sales of government securities, attended by an increase in the volume of bank credit. These new creations of purchasing power, as we know, exert a continued upward pressure on prices.

Here again, speculation plays a very important part. The movement of prices in a period of continuing inflation is not a direct effect of the sheer mechanical impact of the offer of new supplies of money in the market for goods and services. Speculation anticipates these consequences, and, by anticipating them, puts into operation a new mechanism of its own. Except in the initial, and perhaps the very last, stages of inflation, the price of foreign exchange and of imports and exports generally keeps in advance of the general movement of domestic prices, even when external debts are not exerting a steady pressure upon the foreign exchange market. The general level of domestic prices, though again lagging behind, follows the course marked out for it by the appraisal of the value of the country's currencies in the international market. Thus, for example, fresh issues of paper money to cover a deficit in the budget may, merely by altering the expectations people have formed respecting the deflation or stabilization of the currency, lead to a sharp advance of prices, brought about by increased offers of currency for goods or for foreign exchange. This in turn will call for and in a manner justify further creations of bank credit. In a similar way, any really effective steps taken toward deflation

will produce a powerfully cumulative downward influence. Under a régime of inconvertible paper money, it is barely a step from a stable to an extremely unstable equilibrium. So far as I know, all of the world's experience with inconvertible paper money illustrates and confirms these principles.

The bearing of these considerations upon what we may call the international problem of national debts is obvious. The stabilization of European currencies (I do not mean their drastic deflation) is the indispensable prerequisite to the economic recovery of Europe. The one indispensable instrument of stabilization is the balancing of budgets. The burden of debts, external and internal, blocks the way. If the debts can be taken care of, the monetary situation would take care of itself. This is what makes of national debts, seen as an aggregate, an international problem — and no shutting of our eyes to it will alter the fact.

I am not one of those who believe that the payment of the interest and principal of an international debt does more harm than good to the nation receiving the payment. Nor do I believe that the payment of an international debt strengthens rather than weakens the nation making the payment. But I am convinced that the present debts of European nations, including reparations, interallied debts, and internal debts, are so large that, as they now stand on the books, they cannot be paid in full. Moreover, no considerable part of them can be paid without perpetuating and perhaps increasing the disorders of the currency. There is, I grant, one possible alternative. These debts might within the next generation be reduced to manageable proportions by the persistent use of taxation of unprecedented severity. I doubt that this is politically practicable. Even if it

were, I doubt that the game would be worth the candle. There would be long periods of business depression, lowered standards of living, and new injustices in the distribution of wealth. It would be better to release the world's economic energies; to permit deflation to be accomplished, so far as it is desirable, by increasing the volume of production and trade rather than by drastic reductions in the volume of the currency.

On the whole there is less pessimism respecting the general economic condition of Europe than there was a few years ago. Despite the enforced retrogression of Germany during the past year, Europe as a whole continues to go forward and not backward. But her complete recovery will be delayed indefinitely if there are to be either further increases of debt and further inflation, or a steady general reduction of debts and deflation of currencies.

III

HUNGARY IN 1925[1]

"HUNGARY cannot, of course, be in a sound situation until both her budget and financial position and also her trade balance are satisfactory. She must not only meet her public expenditure by taxation but she must produce (and dispose of) as much as she consumes."

I begin with a quotation from the Report of the Financial Committee of the League of Nations (December 20, 1923) because that quotation defines the problem with which the present memorandum is concerned. The primary objects of the plan for financial reconstruction — the stabilizing of the currency and the balancing of the budget — have already been secured, and, in the case of the budget, much sooner than had been expected. The questions that remain relate to the future. Is it likely that before the situation has become sufficiently strong to resist them, strains and shocks will undo some of the work that has been accomplished?

I. POPULATION

I have met with the view that Hungary is overpopulated, that the full restoration of prosperity is impossible until an outlet is found for her surplus population. I do not find that this view is sound. The density of population in Hungary (86 per square kilometer) is fairly high — higher, for example, than in Austria or

[1] A memorandum prepared in Budapest in August, 1925, at the request of Jeremiah Smith, Jr., Esq., then Commissioner General of the League of Nations for Hungary.]

France, although lower than in Czecho-Slovakia, Italy, or Germany. But comparisons of this kind mean little, unless differences in natural resources, in standards of living, and in the composition and distribution of the population are also taken into account.

It is better to compare the present population of Hungary with the pre-war population of the same area. At the end of 1924 there were some 8,200,000 persons in Hungary (estimate of the Central Statistical Office) — about 11 per cent more than were within the same boundaries in 1910. Even if there is some exaggeration in the estimate, there has been a notable rate of increase — higher, probably, than in any other European country which participated in the war. However, the natural increase of the population (the excess of births over deaths) since the war has been no higher than the average for western Europe, and distinctly lower than the rate which prevailed before the war within the present boundaries of Hungary. There has been a marked diminution of the birth-rate, especially in Budapest, where, in the last few years, deaths have been more numerous than births. The general death-rate has also diminished, but merely because young children now constitute a smaller proportion of the population. The decrease of the infantile death-rate has not been so marked as in other parts of Europe.

The rapid growth of the population is attributable to the influx of Hungarians from the regions that were assigned to other states by the Treaty of Trianon. The larger part of this movement, it is safe to assume, has been completed. There is nothing in the resulting situation which affords any basis for pessimism.

In the first place, a larger proportion of the population than formerly is found in the *productive* age groups.

The number less than 15 years of age was smaller in 1920 than in 1910. This is at best a temporary economic advantage. For the next generation it will become a disadvantage. The number of possible productive workers will then be relatively small, as compared with the number of the aged. This, however, is a handicap Hungary will share with most of the other countries of Europe. And the difference, one way or the other, is not a matter of prime importance.

POPULATION OF THE PRESENT TERRITORY OF HUNGARY, BY AGE GROUPS: 1910, 1920

AGE	NUMBERS		PER CENTS	
	1910	1920	1910	1920
Under 15	2,643,745	2,443,555	34.8	30.6
15–60	4,355,778	4,816,754	57.2	60.3
Over 60	607,005	715,990	9.0	9.0

Before the war, the average annual rate of growth of the population of (present) Hungary was higher than it has been in recent years. The only ground for concern, therefore, is that the population has increased while the accumulated capital of the country has been depleted and its economic organization disrupted. It is to be remembered, however, that except where population presses closely upon the means of subsistence, increased numbers are an asset as well as a liability. They increase the possible aggregate national production of wealth, even though, under some conditions, they may decrease somewhat the possible amount of wealth that can be produced per capita. "Surplus population," after all, is generally a purely relative term. Hungarians migrated to America before the war, not because

they could not earn a living at home, but because they could earn a better living in America. The Hungarians, moreover (as the "one-child-family system," prevalent in parts of Hungary, indicates), have shown a capacity to regulate the growth of their population in accordance with economic conditions. Diminished emigration is likely to be compensated for in part by a lower birth-rate. In short, I conclude that overpopulation does not constitute a serious problem, either of the present or of the calculable future.

Doubtless more weight must be given to the contention that the population of the country is badly distributed, and that Budapest is too large a capital for so small a country. In 1920, the cities contained 19 per cent of the population, and towns with organized councils 13 per cent. About 2,400,000 persons, making about 30 per cent of the total, lived in cities and communes having a population of more than 20,000, and nearly 40 per cent of these were in Budapest. As compared with conditions in western Europe, these figures do not reveal a high degree of urban concentration. Here again, external comparisons throw little light on the Hungarian situation. The important thing is that the urban populations of Hungary, and notably the population of Budapest, have been increasing much faster than the rural population during a period in which urban economic life has been more seriously retarded than agriculture has.

There is no one right balance as between the urban and rural population of a country. Other important variables enter into the situation. What Hungary can produce most efficiently, what she can export and import to best advantage, will depend in some measure upon her population and how it is distributed, as well as

upon her "natural" advantages. It is clear, furthermore, that it is the *market area*, rather than the extent of a country's territory, which is important in its relation to the size of that country's cities. Whether Budapest is or is not overpopulated is a question which cannot be answered satisfactorily without taking into account the customs policy of Hungary and her neighbors.

One thing, however, seems fairly certain. As things now are, there is a *relative* excess of urban population. Furthermore, despite the operation of forces tending toward a better adjustment (slackening of the movement from the country to the city, low urban birth-rate, etc.), this relative excess is likely to continue for some years and to have a number of fairly definite effects. Among these effects are: (1) a relatively low level of real wages; (2) a relatively low industrial output per capita; (3) a wasteful multiplication of middlemen; (4) a tendency to rely upon state aid rather than upon productive efforts. I do not mean to imply that in these respects the Hungarian situation is unique or that it is particularly serious, and I do not forget that such things have a way of righting themselves in time. Our present concern, however, is with what, if things go well, will be a transition period, and during that period the matters I have just reviewed promise to have some importance.

II. PRODUCTION

The crops, more than any other single factor, determine the degree of Hungary's prosperity. As an index of her general economic progress, however, the crop yield, variable as it is from season to season, is less significant than the *acreage* devoted to crops. The facts are summarized by the following figures:

AREA DEVOTED TO CROPS IN HUNGARY
(average for 1911–1915 = 100)

	1920–1922 (average)	1923	1924
Grain	87	90	92
Maize, potatoes, beets	102	108	108
All crops	94	94	97

The figures show (1) that there has been steady progress toward the pre-war level, and (2) that there has been some shifting from grain to other crops. The shifting appears to be because of the barriers in the way of the free movement of agricultural produce between the portion of her former territory which Hungary has retained and the portion which she has lost. The territory which she has retained formerly produced about 47 per cent of Hungary's grain, and only 38 per cent of her cultivated crops (maize, potatoes, beets, etc.). The product obtained, however, has not been proportionate to the effort put forth. Taking the yields in both 1923 and 1924 into account (1925 was a better year for grain, 1924 for other crops), and giving some weight to the relative values of the different crops, it appears that the product fell short of the pre-war level by about a fourth. The grain crop for 1925 promises to be much larger.

The diminished yield per acre is attributed by agricultural experts not so much to unfavorable seasons as to a shortage of both natural and artificial fertilizers. The growing of grain in Hungary requires the use of phosphates. The raw material, before the war, came largely in the form of phosphate rock from Florida and Northern Africa. In 1913 about 14,000 carloads were used within the present boundaries of Hungary. After the beginning of the war none was imported until 1922, and in 1924 only 4500 carloads were used. The difficul-

ties have been, first, high costs and adverse exchange, and second, high interest rates and a shortage of credit available for advances to farmers. Other artificial fertilizers, also, have been hard to obtain in sufficient quantities.

Domestic animals are an important element in Hungarian husbandry. At the time of the Treaty of Trianon their number was seriously depleted. Since then there has been a rapid increase. In 1924 the number of sheep and hogs in Hungary was about three fourths of the pre-war number. The number of cattle was 12 per cent and of horses only 5 per cent less than before the war. The figures for 1925, I am informed, will show further gains.

Summarizing the facts which have just been reviewed, it appears that the acreage devoted to crops has been restored to substantially the pre-war level. This means, however, a net drop of about 10 per cent in the ratio of crop acreage to total population. The agricultural output, however, partly because of a shortage of working capital, has lagged behind. It has been not more than three fourths of the pre-war figure. This means that the agricultural product per capita (of the total population) has been only about two thirds as large as before the war. Even the large crop in prospect for 1925 may fall considerably short of the pre-war average. It is probable, however, that with favorable seasons the pre-war output may soon be reached or even surpassed. It will be a more difficult matter, however, to equal the pre-war volume of production per capita.

Turning from the agricultural to the industrial situation, we find it much more difficult to get reliable measures of losses and gains. Statistics are available for a number of the large industries, and further information

has been gained in interviews with representatives of some of those industries. The amount of unemployment and the general level of real wages also throw some light upon the problem. I have given little weight to figures that run in terms of "percentage of capacity," for these are likely to be misleading. Even in the best of years, the total industrial output always falls short of the aggregate rated capacity of a country's industries.

Taking the last half of 1924 and the first half of 1925 into account, it appears that most of the larger industrial establishments, such as depend in considerable part upon export markets, have been producing from 25 per cent to 65 per cent of their pre-war output, the average being, I should say, under rather than over 50 per cent. On the other hand, there are certain industries, notably the textile industry, and in a smaller way the shoe and leather industry, which have made substantial gains. Furthermore, there are many small industrial establishments producing for local markets, whose output, it is safe to assume, fluctuates in fairly close relation with changes in agricultural production. I do not believe I am far wrong in estimating the total industrial output in Hungary during the twelve months under review at about 60 per cent of the annual output just before the war, or, on a per capita basis, at about 54 per cent of the pre-war amount.

After allowance is made for the decrease in the value of gold, this result is not inconsistent with the industrial statistics collected by the Hungarian Central Statistical Office, which report that the *value* of the industrial product in 1924 was 1447 millions of gold crowns, as against 1650 in 1913 (and only 711 in 1921). If the official figures can be trusted, the number of laborers employed in industry was fully as large in 1924 as in

1913. The lower output per laborer probably reflects, first, a general reduction of about a fifth in the length of the working day; second, some decrease in the efficiency of labor (although there is common testimony that this is rapidly improving); third, relative gains on the part of industries to which cheap labor gives especial advantages and which have a comparatively low output per laborer.

The figures for other elements that enter into the national production of wealth (hand trades, mining, trade, transport) need not be examined in detail. Even in the aggregate they are much smaller elements than agriculture and industry. Taken together, however, they have been at a relatively higher level than industry, and possibly at as high a one as agriculture. In making up the composite result, agriculture should be given about four times as much weight as industry, or — putting the minor elements in with agriculture — about five times as much weight. The possible margin of error is too great to justify precisely formulated results. I believe, however, that we are warranted in concluding that the production of wealth in Hungary during the year ending June 30, 1925, was little if any more than two thirds as large as before the war, and that the per capita production of wealth was probably not more than three fifths of the pre-war amount.

III. FOREIGN TRADE AND COMMERCIAL POLICY

Hungary's foreign trade has always been in the nature of border traffic. Before the war, about two thirds of it was with Austria and most of the balance was with neighboring states. To-day a similar situation exists, except that what formerly was domestic trade now passes across the new national frontiers. In 1924 nearly two

thirds of Hungary's trade was with Austria, Czecho-Slovakia and Yugo-Slavia. Add Poland and Germany, and 85 per cent of Hungary's foreign trade is accounted for.

Under normal conditions, the establishing of the new national boundaries should have increased the proportion of foreign to domestic trade. This is not what has happened. Before the war, Hungary's foreign trade, figured on an average of her exports and her imports, amounted to about 30 per cent of her annual production of wealth. At the present time, the corresponding proportion probably is not more than 20 per cent. Hungary's situation in this respect is like that of other countries of Europe, and the factors responsible for the decrease of her foreign trade are mostly such as have affected other countries as well. Currency depreciation and disordered exchanges, despite statements sometimes made to the contrary, diminish rather than increase foreign trade. Export and import prohibitions, new customs frontiers, higher tariffs and, in some cases, higher costs of transport have also played a part. Hungary, whose chief exports are agricultural products, has also been adversely affected by the diminished purchasing power of the more highly industrialized countries of Europe.

The primary factors which determine the value of Hungary's exports are (1) her crops, and (2) the prices that can be got for them. The rapid improvement already made in respect of the first of these factors has been discussed. The second factor is not under Hungary's control. The course of agricultural prices varies from season to season, but the general level in recent years has been low. The level depends in part upon the general balance between agriculture and industry, and

agriculture in Europe generally has come much closer to making a complete recovery than industry has. As compared with many of the countries of western Europe, Hungary is advantageously situated; but her general rate of progress will be governed, in some measure, by that of the rest of Europe.

This leads us to a consideration of Hungary's commercial policy. For a country situated as Hungary is, the power of a customs tariff, either for good or for evil, is not large. Hungary will sell most of her exportable surplus of agricultural products under one policy or another. For years to come, it will make no *large* difference in her balance of payments whether her wheat passes through the mills of Budapest before being exported, whether her wool is spun and woven in Hungarian mills and consumed at home, or whether it is exported to other countries from which Hungary would buy cloth.

The Financial Committee of the League of Nations was undoubtedly right in urging that Hungary should devote her energies to the production of those things which she can produce to best advantage; yet, as I have already suggested, the things which she can produce to best advantage are determined, in part, by other factors than her land and other natural resources. For one thing, the tariff policy of her neighbors must be taken into account. For another thing, her present labor situation has an inevitable influence upon the trend of her economic effort. International comparisons of labor costs are hazardous. There is no doubt, however, but that the present level of money wages in Budapest is extremely low — somewhere near the pre-war gold level and probably under rather than over that level. Wages in Italy and Czecho-Slovakia may be as low as in

Hungary, but in most of the industrial countries of Europe they are distinctly higher.

It is inevitable that efforts should be made to utilize this surplus labor supply, especially as the opinion of employers is that Hungarian workmen are fairly efficient. The rapid development of the textile industry is a case in point. Taking the world as a whole, that industry, since the war, has been decentralized to a considerable extent. Customs barriers and export and import prohibitions have had something to do with its development in many different countries. Another important factor has been a surplus labor supply in countries which were formerly importers of textiles. Unable to produce and to sell their former exports to advantage, countries like Hungary have turned toward producing some of the things they formerly imported. Doubtless there is real waste here; the world has lost, and is losing, some of the advantages of international economic cooperation. Nevertheless, given the situation as it is, the trend toward great national self-sufficiency is inevitable. I mean that it is not merely a matter of the artificial protection created by the war and by the disordered state of affairs which followed. It is, in large part, a result rather than a cause of the stoppage of international trade that has come from disordered exchanges, and most of all from the situation in respect of international debts.

I have not been able to examine the new Hungarian customs tariff in detail. It is clear, however, that the duties it imposes are, on the whole, distinctly high and that it bears more heavily upon cheap than upon expensive goods. Such study as I have been able to make of it, however, leads me to the opinion that it does not embody a general protective policy. Hungary's largest

industries, as well as her farmers, have a larger interest in keeping foreign markets open than in closing the home market. I gather that the new tariff has been designed partly to afford protection to certain industries that have grown up in Hungary since the war and partly to afford a weapon for use in bargaining for tariff concessions from other countries.

Hungary is disadvantageously situated in respect of tariff negotiations. More industrialized than Rumania or Yugo-Slavia, she is, as compared with most of the other countries with which she trades, distinctly an agricultural country. Some of her customers are interested in one type of tariff concession, some in another. Under these conditions, the operation of most-favored-nation clauses is likely to produce a general downward movement of her tariff. In the meanwhile, as I have indicated, there is nothing in her present tariff policy which gives serious cause for alarm. I do not believe that it will exercise a large adverse influence upon her balance of trade, although, of course, it will not have the beneficial effect that some of its advocates appear to expect. It would be vastly better for all concerned if the customs barriers that have been erected within the old Austro-Hungarian empire could be done away with completely; but, that being just now impossible, Hungary's present policy must be judged in its relation to the actual situation.

IV. TAXATION

I shall not attempt to review the whole field of Hungarian taxation. I shall discuss only two matters: (1) the amount of the burden imposed by taxation; and (2) the general character of the tax system.

I begin with the estimate made by Professor F. Von

Fellner of the amount of Hungary's national income before the war.[1] Professor Fellner first estimated the aggregate national income of former Hungary, and then apportioned it among present Hungary and the successor states. The methods he used in making the apportionment were detailed and elaborate. I have some confidence in them, not because they were detailed, but because the share apportioned to present Hungary (about 40 per cent) is about what is indicated by other indexes of the relative economic importance of the territory Hungary retained and the territory she lost. The accuracy of his result depends, I suspect, upon the accuracy of his estimate of the national income of the larger Hungary. Although that estimate was carefully made, any estimate of the sort is likely to have a possible error of as much as 20 per cent, although the probable error is less. Furthermore, Professor Fellner counts as income only the value of the *material goods* produced annually together with the net balance of foreign credit items. He omits the professional and other personal services which are rendered directly within the year, as well as the services of Government. For this reason his results are not properly comparable with the current estimates of the national income of other countries, notably of Great Britain and of the United States.

Professor Fellner's estimate of the aggregate income, just before the war, of the territory constituting present Hungary is close to 3,000,000,000 gold crowns. On the basis of very inadequate evidence (occupation statistics, governmental expenditures) I venture to increase the estimate to 3,600,000,000 gold crowns, so as to include items that were omitted. Even with this modification, I suspect that the estimate is too low rather than too

[1] *Metron*, September, 1923.

high. And a further increase of 40 or 50 per cent must be made in order to take account of the subsequent decrease of the purchasing power of gold.

We have already estimated that the total annual production of wealth in present Hungary is not more than two thirds of what it was before the war. The annual production of wealth, of course, is equal to the aggregate annual income of the country. Because precise results are impossible, we may assume that this shrinkage of one third in the total national produce is offset by a shrinkage of about one third in the value of gold, so that the gold value of the present national income of Hungary is about what it was before the war.

If the total amount of taxes in Hungary is estimated by taking the revenue anticipated by the Budget for 1925–26, deducting that part of the revenue from the tobacco and salt monopolies which could not be fairly construed as taxation, and adding 100 millions on account of municipal and local taxation, we reach a sum of about 600 million gold crowns. This sum amounts to between 16 and 17 per cent of the 3,600,000,000 which we have set down as a rough estimate of the aggregate national income. We are fairly safe in concluding that taxes in Hungary probably absorb between 14 and 18 per cent of the total national income, and quite certainly more than 12 and less than 20 per cent. It is possible that the amount raised by taxes will be considerably larger than the Budget estimates, but this probably will occur only in case the national income is also correspondingly increased.

The burden of taxation depends not only upon a country's wealth and income, but also upon the nature and incidence of its tax system. Where the average income is small, the burden imposed by a given rate of

taxation is greater than where the average income is large. To take 15 per cent of the income of Hungary by taxation is to impose a burden equal to what would be imposed by, say, 20 per cent in France. By almost any standard of comparison, Hungary is heavily taxed, even if not so heavily as some of the other countries which participated in the war.

Hungary's chief reliance is upon indirect taxes. As compared with direct taxes, indirect taxes have been in bad favor among experts in public finance, although, I think, they are somewhat more favorably regarded than they were twenty years ago. Their burden is greater than that imposed by direct taxes, in that they bear relatively more heavily upon the poor than upon the rich, although indirect taxes which have remained at a fairly constant level for a number of years are likely to have their burden pretty well diffused. Their repressive effect upon trade and industry is less than that of direct taxes.

I see no reason why Hungary should hasten to make any radical changes in its present revenue system. An effective progressive income tax ought to be installed as a partial offset to the indirect taxes. I do not believe that much can be accomplished by attempting to improve the administration of the present income tax. A new tax framed on new lines would be much more likely to succeed. The success of income taxation depends very largely on securing the coöperation of the taxpayer. This is a reason why it is useless to try to revamp the present income tax, which is now pretty thoroughly discredited. It is also a reason why the rates initially imposed, and the degree of progression in the rates, should be moderate.

V. THE BALANCE OF PAYMENTS

I have given much of my time during the past three weeks to the study both of the details and of the general aspects of this problem. In this memorandum, however, I shall confine myself largely to general aspects and shall give small attention to details. This for two reasons: First, an analysis of the balance of payments has been made by Dr. M. Frère. He and I have had access to the same sources of information and we have discussed them together. The results I should get would not differ substantially from his and certainly would be no more dependable. Second, there is less doubt about the facts than about what the facts signify. The adverse balance of commodity trade, the rapid increase of the reserves of the National Bank, and the relatively small size of most of the directly measurable invisible items are sufficient in themselves to establish the general character of the changes that have taken place during the past year. The net result is that the reserves (gold and foreign exchange) of the National Bank have increased pari passu with the increase of the country's funded and unfunded international indebtedness. How sound is the situation thus created?

One contention which is fairly sure to be advanced can be summarily disposed of. That contention is, in effect, that the situation is inherently unsound merely because the Bank's reserve position is the outcome of borrowing rather than of a favorable balance of commodity trade. According to this view, if next year's balance of trade should be heavily adverse, disaster would ensue. Called upon to meet the deficit and to repay the short-time loans then outstanding, the debtors would drain the Bank of its reserves, and it would no longer be possible to keep the currency stabilized.

This analysis is in error because it fails to take account of two important considerations: (1) Whether the foreign credits extended to Hungarian borrowers would be curtailed or extended does not depend upon the balance of its international indebtedness at any given time, but upon the *outlook for the future*. (2) The balance of commodity trade is not a wholly independent variable, determining the increase or decrease of the amount of a country's foreign indebtedness. The truth is that international borrowing and lending in their turn have a potent influence upon the balance of commodity trade.

This influence is exerted in two different ways. First, the proceeds of foreign borrowings may be expended for foreign goods, thus immediately increasing imports. Such transactions are sound when (*a*) seasonal changes or other causes may be expected later to bring about a compensatory increase of exports, or when (*b*) the imports bought with the borrowed funds are of such a character as to increase the country's capacity to produce and export goods.

Second, the proceeds of foreign borrowings may be expended within the borrowing country, the foreign credit being retained as an asset (as, for example, in the reserves of the central bank, which gives domestic currency in exchange for it). In this case, there is no immediate increase of imports. Instead, the sequence of events, under some conditions, may be as follows: The expenditure of the new purchasing power within the country increases the demand for goods and services. An upward pressure is exerted upon prices. Business activity is stimulated. Profits increase. More loans are secured abroad. The process continues in a cumulative way until one or more of a number of possible causes bring it to an end. Business depression and loss

of confidence ensue. The country is called upon to liquidate its maturing foreign indebtedness. The reserves of the central bank are depleted and the currency is endangered. Very likely, the high prices will have attracted imports and retarded exports, thus making the situation worse than otherwise it would have been.

The sequence of events which has just been outlined is a fairly common result of the expansion of foreign borrowings, when the proceeds are expended within the borrowing country. It is not, however, a *necessary* sequence. Much depends upon (1) the degree in which the borrowing country's economic and financial structure is sensitive to inflationary influences; (2) the amount of such borrowing and the rate at which it is increased; and (3) whether the proceeds are used in such a way as to increase the country's capacity to produce and export goods.

Under existing conditions I believe it inevitable that Hungary should increase her international indebtedness. In fact, it is desirable that she should, provided that it be done slowly and steadily. Hungary, like many other countries of Europe, is hampered by a shortage of working capital (I shall not stop to define that much-abused term), and her complete economic recovery would be indefinitely delayed if, on top of her other burdens, she had to accumulate, by saving, all of the capital she needs.

From 1900 on, the balance of trade of pre-war Hungary turned gradually from a favorable to an increasingly unfavorable position, evidence of an increasing investment of foreign capital. The condition thus created was not unsound, and it is to be expected that the same trend should now be taken up anew. Hungary's position in this respect is stronger than that of

countries which before the war were lenders of capital, and are now forced to become borrowers.

Returning now, after these general considerations, to the immediate problem, it is to be observed first that even a favorable balance of commodity trade during the coming year might not diminish the amount of foreign borrowing. Instead, by creating confidence in Hungary's economic future, it probably might increase both the demand for and the supply of foreign loans.

I have suggested, however, that the danger lies, not in borrowing, but in borrowing for non-productive purposes. My judgment is that, so far as these loans are for commercial purposes, negotiated by the industries or the banks, there is no great danger that they will become overextended. For one thing, it is not likely that any very large sums of money are awaiting investment in Hungary. Furthermore, I gather from representatives of important industries that they are more interested in getting markets than in getting credit. I have found no suggestion of a speculative spirit, or any disposition to count upon a general revival of business activity. Nor do I believe that the general structure of industry and finance in Hungary is such that a release of the pressure which falling prices have exerted during the past year, coupled with access to larger supplies of loanable funds, is likely to bring about a strongly marked upward movement of either security or commodity prices. The Budapest stock exchange, for example, has a distinctly local character, and the volume of transactions upon it is not large. Nevertheless, of course, the situation is one which calls for careful watching. The National Bank is naturally fully aware of the possible danger and of the necessity of being prepared to meet it.

The real danger, as I see it, is from such borrowing as

might be undertaken, not in order to reap business profits, but to "better the situation." Projects for securing money abroad for long-time agricultural credits are now on foot. Now long-time credits to facilitate the buying of land are precisely the sort of thing that Hungary does not need. *Transfers* of land do not increase the country's productive capacity. The real effect of such loans depends upon what the *sellers* do with the money. Their net effect, for example, might be increased consumption instead of increased production. Doubtless long-time credit, secured by land mortgage, would sometimes be used to replenish a farmer's working capital. But it is better that credit which is to be used for such purposes should have its maturity and its other terms determined accordingly. There are dangers also, of course, in expenditures which the Government (if released from control) might make out of borrowed funds, either in response to political pressure or from mistaken notions respecting the effects such expenditures have.

VI. THE CURRENCY

In most respects the effects of the successful stabilization of the currency have been such as would be expected. It should be observed, however, that a safe middle course between excessive depression and dangerous speculative expansion was followed with more success than in other countries where stabilization has been achieved.

There were the customary rapid expansion of circulation (so as to overtake the price level), the prompt advance of wholesale prices, the lagging advance of retail prices and wages, the rapid growth of bank deposits as confidence in the currency was restored, and the other

familiar phenomena. The prices of goods of foreign origin, in May of 1924, were very much higher, relatively, than the prices of goods of Hungarian origin, for the price of foreign goods had advanced along with the price of exchange. After stabilization, the prices of foreign goods remained fairly steady (except for a sag late in 1924), while the prices of domestic goods advanced, although not steadily, until early in 1925 when a downward movement began, which was soon shown in the prices of foreign goods as well.

Some part, of course, of the characteristic movement of the prices of Hungarian goods is attributable to the movement of the world prices of agricultural products. A considerable part of it, however, was brought about by stabilization, which permitted the prices of the two types of goods to come into better alignment. If it had not been for this price movement, favorable to Hungary, the adverse balance of commodity trade for the year ending June 30, 1925, would have been considerably larger than it proved to be.

The general character of the movement of prices during the year, however, was determined by the fact that the crown was tied to sterling rather than to gold. This proved to be a fortunate circumstance. An adequate degree of stability was assured, and yet, with the upward movement of sterling in relation to gold, there was a repressive influence upon any expansive tendencies. The general movements of prices in Hungary and Great Britain since the middle of 1924 have been strikingly alike. The repressive influence exerted by the downward movement of prices was not excessive. Taking the relatively small grain crop into account, the net recession of industrial activity during the year, as revealed by the available indexes, was small. (For the

past twelve months the fluctuations of railway traffic and of employment, which tell much the same story, constitute probably the best available indexes of economic conditions.)

The movements of Hungarian prices in the near future will be determined in part by general changes in the world level of prices, in part by the special forces affecting the (world) level of agricultural prices, and in part by the expansion or contraction of credit within Hungary. I do not expect that this last factor will play an important part. The character of the movements of prices in Hungary during the past year, as well as the generally solid character of the industrial position, give ground for the opinion that inflationary influences, even if any should appear, would have no large effect.

It is probable, however, that the Hungarian price level is not yet fully adjusted to the general world level of (gold) prices. Comparisons with other countries, it is true, indicate that in respect of the net increase of wholesale prices since 1913, Hungary occupies something like a median position. The index numbers of different countries, however, utilize different lists of commodities, are differently constructed, and are only imperfectly comparable. The Hungarian index number (published by the Central Office of Statistics) is, for example, an unweighted arithmetic average of relative prices. Index numbers so constructed are known to have an upward bias — that is, they give an exaggerated report of the percentage increase of prices. If this were its only doubtful point, we could be fairly sure that the Hungarian index number is too high by from three to ten points. There is a chance, however, that other peculiarities may offset or more than offset its upward bias.

An adequate analysis of the present position of the National Bank would call for a good deal of technical detail. I shall limit myself to a bare summary of some important points.

1. The Bank has been ably and prudently managed, but its contact with the market is neither so broad nor as firmly established as could be desired.

2. The wide margin between the Bank's rate of discount and the short-time rates prevailing elsewhere in Hungary suggests that the Bank's rate is ineffective, and that it has little or no control of the credit situation. In fact, however, the Bank probably has more control of the situation than appears on the surface. The bank rate applies to a higher grade of transactions than most of those to which the other rates pertain. Moreover, although not all of the largest commercial banks are borrowers at the National Bank, yet through private deposits and interbank payments the notes issued by the Bank are distributed among all the banks, and all would feel the pressure if discounting were curtailed.

3. The Bank has virtually no control, however, over the amount of notes put into circulation by the purchase of foreign exchange. So far as such exchange originates in foreign borrowing the Bank acts merely as money-changer. The real portfolios are held abroad, and the discount policy is determined there. Nor is the Bank responsible for an increase of circulation resulting from any financial operations on the part of the Government.

4. Some of my reasons for thinking that there is no large element of danger in the increase of the note issue by means of foreign borrowing have been given in discussing the effects such borrowings have upon the country's balance of payments. Further reasons are: (i) An advance of discount rate would be likely to dimin-

ish the *aggregate* amount of borrowing, and hence to contract the circulation, even if it should increase the amount of borrowing done abroad. (ii) The expansion of the circulation through foreign borrowing is not dangerous unless it leads to a rise of prices. A rise of prices, however, would almost certainly be attended by increased demands for discounts at the National Bank, and this would give it an opportunity to exercise a restraining influence. The Bank's power to act quickly and efficaciously, however, is hampered by its inability to act *on its own initiative* in such a way as to increase or decrease its circulation. It should be empowered to buy and sell securities in such a way as to increase or decrease the burden borne by the other banks.

5. The rigid stipulation that the Government shall not borrow from the bank has some undesirable consequences. (i) If control were withdrawn, it might be possible for the Government to secure small short-time advances from *other* banks to meet a temporary deficit in its cash position, such as might be brought about by seasonal factors. It is very much better that any such dealings should be with the National Bank. (ii) If the Government were permitted to discount treasury bills, limited to a definite percentage of its ordinary revenues during the preceding year, the Bank could use such securities, in a limited way, in operations which would strengthen its control of the credit situation.

6. The most difficult and delicate problem confronting the Bank is that of its discount rate policy. I have given reasons for thinking that the Bank has rather more control of the situation than appears on the surface. In addition, there are the familiar psychological effects of changes of the discount rate. Nevertheless, the terms upon which foreign loans can be obtained by

the other banks and by large industries limit the National Bank's discretionary range. Its discount rate might easily be made so high as to be largely ineffective. It is wiser policy, I should say, for the Bank to endeavor to maintain the right ratio between discounting and foreign borrowing — regarding not only a rapid decrease, but also an unduly rapid increase of its reserves as a sign of danger. In addition to using its own reserve position as a guide, the Bank, in my opinion, should endeavor to secure reliable statistical indexes of changes in the physical volume of production and trade in Hungary, so that it may have some knowledge of the amount of currency which the country needs.

IV

THE TREND OF PRICES[1]

BUSINESS men outside of a few favored industries seem to feel that on the whole the year 1922 has been disappointing. They are looking forward with uncertainty, not unmixed with pessimism. On the other hand, I infer that students of business cycles see nothing abnormal and no large measure of uncertainty in the present situation. Following upon the collapse of 1920, liquidation and recovery seem to have taken their normal course. During the last few months there have been some small but apparently trustworthy signs of the beginnings of another period of credit expansion, and there appear to be no immediate obstacles large enough to check it.

There are some few dissenters who emphasize the undoubted fact that it was no ordinary business cycle that came to an end in 1920; and who argue that in proportion as inflation and the rise of prices were on a magnificent scale, so the period of deflation and of falling prices will be prolonged. A few even hold that prices must return to the pre-war level before they again increase noticeably.

This matter of a possible return to pre-war prices deserves some passing attention. So far as the immediate future is concerned, such a return is not only undesirable but impossible. The rapid rise of prices was attended with economic injustice, unevenly distributed. But

[1 Reprinted, slightly abridged, from Papers and Proceedings of the Twenty-Fifth Annual Meeting of the American Economic Association, *American Economic Review*, vol. XIII, no. 1, Supplement (March, 1923).]

drastic deflation would bring with it a new series of burdens and injustices, only in small part compensating for the old ones. It would be a new instance of two wrongs failing to make a right. Deflation of that drastic sort, moreover, is beyond the powers of those who direct the policies of the Federal Reserve System — and I do not mean that I think there is any reason to suppose that it is what they desire or intend.

Half, or more than half, of the present volume of deposits in the banks of the country has been created by purchases of government and corporation securities, or by loans in which such securities figure as collateral. That much at least of the present volume of purchasing power is inelastic, so far as contraction is concerned. It may be increased, but it cannot be decreased except by the slow processes of taxation and of the accumulation of real savings in advance of the year-to-year requirements of industry. Complete deflation is impossible, therefore, so long as these hardy fruits of inflation yet live.

The general level of prices, declining from its post-war peak, has reached a broad plateau, of which the eye can as yet see no ending. There will be cyclical and other oscillations, but these do not determine the long-time trend. If we turn to past experience for light on the distant course of that trend, we find it of little help. The rise of prices during the Civil War was hardly more than a billow on the surface of the general current of world prices. It was quickly flattened out and drawn into the general downward course of that current. The world-wide rise of prices in the Napoleonic period affords a better parallel. From the peak they reached in 1814 prices declined, at first rapidly and then more slowly, interrupted only by short periods of expansion,

like that of the thirties, until, in the middle of the century, the flood of new gold turned them upward again. Should we infer that in this particular history is likely to repeat itself? Not unless we have no reason to suppose that there are no essential differences in the other characteristics of the post-war periods we are comparing. I believe that there are differences so important that we cannot afford to overlook them.

The thirty years following the Napoleonic period, like the thirty years following the Civil War, were marked not only by a diminished production of the precious metals, but also by the development of means of communication and transport, by the settlement of new and rich territories, by an increase in the rate of exploitation of the world's stores of mechanical energy, and by a rapid expansion of the world's population. What appears like deflation was in no small part an expansion of the physical volume of trade.

Can one look forward to-day and say with any reasoned conviction that a notable further increase in the rate of acceleration of the world's economic progress is probable? — such an increase as would of itself accomplish a substantial part of the work of deflation? Is not the outlook with respect to the increase of foods, raw materials, and sources of mechanical energy rather such as to lead us to contemplate the day when diminishing returns will cease to appear apologetically in our textbooks as a mere tendency, once of some local importance in the England of Ricardo, but since submerged by larger forces? No one could hold that the physical product is not to continue to increase, but is it not likely that its increase will be at a diminishing rate? And is not this probability — or possibility if you will — enough in itself to suggest that we should be slow to assume a

parallel between the next thirty years and the periods that followed the Napoleonic and Civil Wars?

The future of the gold supply, of course, ought to enter into our calculations. The uncertain factors, however, so greatly outweigh the discoverable ones, that the best that could be offered would be mere conjecture. Furthermore, except for times of extraordinary changes in output, like those at the middle and the end of the last century, the gold supply has more the nature of a dependent than of an independent variable. There may be a new California or a new Rand. Failing that — and the world's mineral resources have been so thoroughly mapped that we know the chances are against it — it seems improbable that the world's annual gold output will return to its pre-war level so long as prices remain as high as they now are.

The pre-war production of gold was supporting an increasing price level. The present annual output of $350,000,000 would support a stationary price level, but hardly one so high as we have now. The diminished output of gold, therefore, is likely to exert a downward pull on prices. Offsetting in some measure this downward pull, there will undoubtedly be increased economies in the monetary uses of gold. A given amount of gold will support a larger measure of purchasing power than it did before the war. A quantitative estimate of these opposed tendencies must be a matter of impressions rather than of exact analysis. My own impression — and it is no more than that — is that, even with a declining rate of increase of physical production, something like equilibrium between the output of goods and the output of gold will ultimately be reached only at a distinctly lower price level than is in immediate prospect.

All this has very little to do with prices in the year

1923, except so far as it eliminates some questions and thus narrows the field of discussion. Our immediate interest is with a short period of time — the field of the "business cycle." But the business cycle is generally conceived as a fluctuation within or around a long-time trend. We have made some progress in our analysis if we may conclude that there is little reason to expect that during the next few years the *trend* will be downward.

Two unusual factors in the present situation, I suppose, are in the minds of those who distrust present forecasts based on the sequence of events in past business cycles. These factors are, first, the unprecedented size of our gold reserves, coupled with the facilities the Federal Reserve System affords for the pyramiding of credit, and, second, Europe's uncertainties.

These great stores of gold are naturally a ground of uneasiness. Yet I do not share the fears of those who believe that the revival of business will feed upon those reserves until they are all but completely absorbed in a prodigious rise of prices. This negative conclusion does not depend upon the probability that long before expansion had gone far the Federal Reserve Banks would be in duty bound to do all that they could to restrain it. If such action on their part should come, I do not doubt that it would be effective. But drastic action, I believe, is not likely to be needed.

I do not mean that a plentiful supply of available credit does not lead to business expansion. Some few economists hold that an expansion of credit is an effect, not a cause, of an expansion of business and of an increase in the volume of goods produced. I believe that view to be misleading (rather than untenable) first, because surplus reserves are a necessary *condition* of business expansion, and secondly, because there are certain

relations of a quasi-mechanical sort between low interest rates and business revival.

The most important relation of this sort is found early in the period of revival, before liquidation of the debts left over from the preceding business cycle has been completed. Low interest rates increase the prices of bonds and create conditions favorable for long-time borrowing. Whether short-time borrowing is more profitable at one time than another depends on the differential between profits and interest. Of the two the rate of profits, merely because it is subject to a wider range of variation, may be said to be the more important determining factor. We place the emphasis rather better, perhaps, when we say that a high rate of profits creates a demand for (short-time) credit than when we say that a low rate of interest in the short-time money market attracts borrowers.

But business projects looking farther ahead, and calling for the placing of investment securities, are not determined wholly or even primarily by the current level of profits, especially at a time when the recovery of business from its last collapse is not yet complete. Low interest rates and high bond prices, rather than profits and the volume of business immediately in prospect, are then, in a real sense, the determining factors.

According to the compilations of the *Commercial and Financial Chronicle*, over a billion dollars' worth of corporation securities were issued in the second quarter of 1922 — equaling the amount issued in the last *half* of 1921. Bond prices reached their peak in September, about a month after interest rates had turned upward. Bond issues have since declined, and relatively to stocks, which reflect profits rather than interest rates, they are likely to decline further.

The revival of the building and construction industries is thus connected, in a quasi-mechanical way, with the condition of the money market. And of course there are some marginal business operations which will be undertaken when interest rates are low that distinctly higher rates might prohibit. But, for the most part, once business revival is under way, it may well be that the money market plays a passive rôle, until such time as declining profits encounter a rising rate of interest, and expansion comes to an end.

The forces at work within a period of business expansion, it is the general opinion of competent students, are partly of a mechanical and partly of a psychological order. Expansion is a bull movement in business, reproducing on a larger scale many of the features of a bull movement on the stock exchange. Its dominant feature is the optimistic discounting of income from future sales. For a while optimistic expectations justify themselves. Money paid out in the expenses of production comes back in the form of consumers' demand for goods. And in a period of rising prices, consumers, it is to be observed, go into the bull side of the market on their own account. Larger sales to them at higher prices demand and seem to justify larger advance of bank credit. Again sales increase as consumers' incomes increase, and so the process of expansion continues in a cumulative way until something brings it to an end.

Now our present interest is in *what brings a period of expansion to an end.* If there are no limits except those set by bank reserves we may well view the prospect with dismay, and ask ourselves just how far prices would rise before our present reserve ratio of approximately twice the normal legal limit would be cut in half.

Such an outcome would be consistent with the views of those who interpret business cycles wholly in terms of the money market. Prosperity comes to an end, they hold, because, and merely because, bank reserves reach their end. The growth of deposits diminishes the reserve ratio; money is drained from reserves into hand-to-hand circulation as wages and retail prices increase; finally the higher level of prices leads to an unfavorable balance of trade, foreign exchange rates rise, gold is exported, and the banks are forced to contract credit and thus precipitate a crisis.

Such, in rough outline, is what we may call the banking theory of crises. It is not to be doubted that it describes with substantial accuracy *some* aspects of *some* business cycles. But it is not a satisfactory general account. We have had some crises, notably that of 1893, in which gold exports played an important part. We have had others, like that of 1907, with which gold exports had little or nothing to do. It is clear, furthermore, that transitions from favorable to unfavorable balances of trade cannot explain *world-wide* depressions of business.

If this banking theory of business cycles were wholly adequate, the surest and simplest way of doing away with crises would be to abolish the gold standard. With irredeemable paper currency there should be, if this view is right, no end to expansion and to prosperity. Europe's present monetary situation would be one to be envied. But the fact is, as we know, that no increase of the currency, no increase of credit, will push business and industry up with it beyond a certain point. Resistance or drag grows at an increasing rate. No amount of inflation can do more than delay the inevitable collapse.

We may conclude, therefore, that it is not necessary that a period of business expansion, once under way, should continue until the surplus lending power of the banks of the country has been wholly absorbed. Other obstacles might and probably would first retard and then stop it before it had reached that point.

Professor Mitchell, in his careful study of the phenomena of business cycles, found it necessary to attach fundamental importance to the way in which the costs of doing business increase in a period of rising prices. While profits, or expected profits, are increasing, business men compete with one another for labor and raw materials. The supply of each is elastic only within limits, so they have to pay more for them. There is common testimony, furthermore, that as labor becomes more fully employed it becomes, on the average, less efficient. And as the capacity of the railways and industrial plants of the country becomes overtaxed they too become *relatively* less efficient. That is, they are operated at increased expense per unit of output. Interest rates also increase, but this is a factor of a different type, and on the whole of less importance. Interest rates seem again, as in an earlier stage of the business cycle, to have more importance for the long-time than for the short-time money market. High rates make producers reluctant to build new plants, even when the capacity of their old plants has been reached.

Such are some of the principal ways in which costs increase during a period of prosperity. Important as these factors are, however, they do not *by themselves* suffice to bring about a collapse. For the expenses of production, viewed from another angle, are money incomes, and must in turn be expended for consumers' or producers' goods. Demand for goods in general

must grow as fast or nearly as fast as business outlays grow. Why, then, should expansion stop, unless the supply of credit gives out?

The explanation must be found, not in a difference between aggregate demand and aggregate supply, but in maladjustments of demand and supply. As prices increase the distribution as well as the amount of money incomes changes, and hence the incidence of the demand for different types of consumption and production on goods changes. Without a careful statistical study of the problem it would be hazardous to generalize respecting the precise way in which rising prices shift demand. Fuller employment makes labor's relative share larger, while the lag of wages behind prices reduces it. How much farmers can buy depends in part upon the crops and in part upon world demand. It is clear, however, that in a period of rising prices the demand for luxuries must increase faster than the demand for necessaries. It is also clear that prosperous industries not only attract larger investments, but, just because they are prosperous, have larger funds at their own command. Business surpluses grow faster than the general fund of disposable savings.

These few examples must suffice. Overproduction and underconsumption theories miss the point. A period of rising prices is attended with both overproduction *and* underproduction, with underconsumption *and* overconsumption. The expansion of production does not and cannot shift its direction fast enough to keep pace with the changing distribution of demand. In fact it advances under its own momentum in such a way as to increase its unfitness to meet the shifting of buyers' demands. In this way strains accumulate in the industrial system which of themselves would bring

about its collapse, whether in the course of an ordinary business cycle or in a period of paper money inflation.

With surplus bank reserves as small as they were in the United States before the war, and as likely to be absorbed early in the period of recovery by an increase of investment purchases by the banks and by loans on stock-exchange collateral, the supply of credit was undoubtedly an important limiting factor. But the experience of other countries, and notably of France, shows that it is possible for central banks to maintain large surplus reserves without making use of drastic changes of discount rates. It is important to note, however, that these large reserves could hardly have been preserved if the banks holding them had yielded to the pressure of inflationists and had reduced rates in order to delay or ward off a period of business depression. Doubtless under American conditions variations of reserves are bound to be much larger than they were in France; but nevertheless I think it improbable that in an average business cycle the expansion of production and the rise of prices would go far enough to exhaust reserves as large as we now hold, unless inflationists should get control of the Federal Reserve Banks.

There seems to be a general agreement among those who have compared the happenings of the last year with the sequences of events in earlier business cycles that the slow advance of wholesale prices which has continued, with some slight interruptions, during the past year will persist during part but probably not all of the coming year. There is difference of opinion with respect to just when the upward movement will end. I attach much weight to Professor Persons's carefully considered opinion that prices are likely to continue to

rise well into the second half of 1923, possibly through most of the year.

So far as I know no one has ventured to predict the probable extent of the rise. During the past year the increase of wholesale prices, on the average, has been from 12 to 15 per cent. Along with this increase there has been a notable gain in manufacturing output, although we are not yet back to a normal level and are far short of the post-war peak, which according to Professor Day's index, was reached in March, 1920. The chances are that in 1923 the expansion of credit will have a relatively decreased effect on output — even though production is likely to be considerably increased — and a relatively larger effect on prices.[1]

It is difficult to allow for the influence of probable changes in the prices of agricultural products. Despite recent recoveries, the prices of farm products remain low as compared with the general price level. It is possible, although not at all certain, that they will move rather sharply upward. But in any event it is not probable that during the coming year they will fully regain their relative position as compared with the prices of manufactured goods.

Some may say that there are elements in the present business situation which will retard or even prevent a further rise of prices. The banks carried business houses through the collapse of 1920 in a way that was impossible before the Federal Reserve System was created. There have not been the drastic reductions of capitalization that followed some of the earlier crises. Exact comparisons with earlier periods are impossible, but the lists of the corporation bonds put on the market during the present year give the impression that an

[1] This forecast proved to be incorrect.]

unusual proportion of them have been issued to fund floating debt or to provide working capital. Business houses, it might be inferred, are not on the whole in so sound a condition as in corresponding stages of earlier cycles.

Even if this is true — and I suspect the contrast with earlier periods may be overdrawn — I cannot see that it has much bearing upon the immediate course of prices. And at any rate there are compensating factors. Thanks in part to the effects of federal taxation upon business policies, the net amount by which business establishments are overcapitalized is probably now unusually small.

The European situation is a highly uncertain variable. If Europe should merely continue on the way of slow but real improvement under the surface of political and financial turmoil, the course of prices in this country would hardly be affected; for our present price structure may be said to be adjusted to the present European situation. The collapse of the paper currency of Germany or of any other important country of western Europe, accompanied by a writing down of its internal debts, would, after the immediate reaction had passed, *increase* Europe's ability to borrow from us. The placing of any large European loans in this country would lead to increased exports and would make prices in this country correspondingly higher and our reserve ratios correspondingly lower. This possibility, however, appears to be contingent upon the willingness of the reparation claimants to waive the first liens they now hold on Germany's power to pay.

Thus far I have assumed a purely passive attitude on the part of the Federal Reserve Banks. I believe that slight advances in rediscount rates, largely by reason

of their psychological effect, would retard the expansion of business and diminish the rise of prices. More drastic steps would be less likely to hold things where they are now than to force a reaction in prices. If prices are permitted to take their normal upward course, and reaction comes later, the Federal Reserve Banks, if they still hold surplus reserves, are likely to be pressed to use inflationary measures to postpone the inevitable.

The economic and political strength of the Federal Reserve System is fairly sure to encounter a severe test before long. Maladjustments in the price structure and the uneven progress of different industries make it hazardous as well as unwise for the Federal Reserve Banks to act now. On the other hand, a purely passive policy will in the long run only accentuate the difficulty of their problems. Somewhere between the two extremes of premature action and of action too long delayed they may find, it is to be hoped, a right middle course.

V

THE STRUCTURE AND POLICIES OF THE FEDERAL RESERVE SYSTEM[1]

I. THE BASIC PRINCIPLES OF CENTRAL BANKING

A GOOD deal of wisdom went into the planning of the Federal Reserve System. With its semi-autonomous regional banks and their branches and agencies, and with its supervising board at Washington, it fits extraordinarily well into the economic geography of the country and into its general political framework. The wisdom of the plan, however, is in these large structural features, not in its detailed specifications. Within very wide limits the details might have been other than they are without materially affecting the way in which the system operates. The actual working organization of central banking in the United States is bound to be the result of slow growth, of gradual adjustment to the requirements of the industrial and financial life of the country. In this process of growth and adjustment there will necessarily be departures from the probable "intent" of the authors of the Federal Reserve Act. An analysis of the intent of the authors of the Federal Constitution would give a very inadequate picture of the actual working of constitutional government in the United States to-day. The Federal Reserve Act, in its important aspects, is no more than the written constitution of our central banking system.

The fundamental principle of central banking is very simple. It is merely that there should be a bank or group of banks set apart from ordinary commercial

[1] Reprinted from *The Annalist*, May 6 and May 13, 1927.]

banks by the circumstance that their policies are shaped with an eye not merely to their own profits and their own solvency, but to the continued solvency of the banking system of the country, taken as a whole. They do not rely upon other banks or upon the market to take paper or securities off their hands in case of need and to supply them with cash. Instead, they always stand ready to perform that service for other banks. For this reason they have to hold among their assets what to a competitive bank would be a disproportionate amount of idle gold and currency. They must always be in a position to meet both export gold drains and variations in the country's requirements for money to be used in making hand-to-hand payments.

A bank which assumes these responsibilities and meets them successfully is a central bank. In order that it may operate successfully, a central bank needs certain powers. It does not, however, have to have all the powers which the Federal Reserve System has. For example, it is not at all necessary that other banks should be compelled to keep the amount of their deposits in the central bank constantly equal to any fixed proportion of their own deposit liabilities. It is not even essential — although it is in many ways convenient and otherwise desirable — that other banks should ordinarily have any funds on deposit in the central bank. It is not necessary, particularly in countries where large use is made of checks, that central banks should have the power to issue notes. But if a central bank is to operate effectively, it is necessary that other banks in the same country should not have an unrestricted and elastic power of note issue.

A central bank, stripped to its barest essentials, is a bank which is always ready to absorb and hold tempo-

rary surpluses in the country's stock of gold and currency, and which always holds itself ready also to supply gold and currency from its own stores upon demand. The other functions which a central bank ordinarily has follow naturally from the responsibilities it assumes as custodian of a country's (or a region's) idle cash. Deposit accounts with it serve perfectly well as reserves for other banks, and hence as means of paying clearing and collection balances. An elastic power of note issue is of advantage to it, because that power enables it to meet variations in the demand for currency without dipping too deeply into its own reserves.

No central bank could assume its necessary responsibilities and pursue a purely passive policy. If it did, sooner or later its reserves would be drained dry by gold exports or by a flow of currency into hand-to-hand circulation. There are two ways in which it can protect itself. First, by advancing its discount rate it can alter the terms upon which rights to demand money from it can be obtained. Such action is likely to have a tempering effect upon the demand for these rights, and will generally shift some of the demand to other markets. Second, a central bank can diminish its cash liabilities to other banks or, under some conditions, increase its holdings of money, by selling some of its earning assets. Other banks will generally find it necessary to supply themselves anew with an adequate amount of claims upon the central bank, but they can do so only upon the central bank's own terms. Under different conditions, of course, the central bank may safely adopt the opposite policy, reducing its rates and increasing the amount of its cash liabilities, so as to permit a general expansion of credit.

Even if a central bank's operations are guided only

with a view to making sure that it will always be able to supply gold or currency on demand without enlisting the help of other banks (and such, before the war, was the professed policy of the Bank of England), it will generally have some stabilizing effect upon the general course of trade and industry. Since the war new importance has been attached to the effects which the operations of central banks have on the general condition of business, and it is quite generally held that their policies should be determined with primary reference to the securing of the maximum practicable degree of business stability.

There are some who deprecate any interference with "the natural course of business." They forget that the operations of central banks cannot but have their effects on business, and that it would be blindly stupid not to take these effects into account when determining just what a central bank shall do and when it shall do it. There are others who object to putting such enormous powers into the hands of the men who control the central banks. No institutions, they say, should be permitted to interfere in an arbitrary way with the fortunes of the business community.

What these critics fail to see is that these are not separable powers which might be granted to or taken away from the central banks. They are bound up with the responsibilities which central banks assume, and cannot be separated from them. Any bank — it need not be a central bank — which is willing to forego some of its profits can operate upon the market "arbitrarily" up to the limits of its resources — as, for example, by selling securities and accumulating gold and currency. Ordinary commercial banks prefer profits to the responsibilities which go with this kind of power. They do not

care to maintain idle surplus reserves. Central banks make it their business to hold surplus reserves. They are usually in a position, therefore, to give other banks access to these reserves on easier terms without concerning themselves about the effect on their own solvency, and they are always in a position to make such access more difficult without concerning themselves about the effect on their profits. The fundamental responsibility of a central bank is to maintain surplus reserves. Its surplus reserves, in turn, coupled with its willingness to forego profits if need be, are the foundations of its power.

The power which central banks have, therefore, is not a privilege "conferred" upon them. It is inseparably bound up with the duties and responsibilities which they have to assume. Only since the war, however, have central banks come to realize how much depends upon the right use of their powers and how difficult it is to know just how and when to use them. No one really knows very much about the deeper-reaching, as distinguished from the merely surface effects of open-market operations and changes of discount rates. Nobody knows just how much or how little the operations of the Federal Reserve Banks have contributed to the general stability of industry in recent years. Nobody knows what the surest symptoms are of an approaching expansion or recession of business activity. No one is in a position to say with any assurance just what the specific criteria are which should guide the policies of the Federal Reserve System.

In fact, we can be certain that reliance upon any simple rule or set of rules would be dangerous. Economic situations are never twice alike. They are compounded of different elements — foreign and domestic, agricultural and industrial, monetary and non-mone-

tary, psychological and physical — and these various elements are combined in constantly shifting proportions.

"Scientific" analysis, unaided, can never carry the inquirer to the heart of an economic situation. Judgment and wisdom — the power to take a complex set of considerations into account and come to a balanced view of them — are quite as much needed as facts and theories. The Federal Reserve Banks need to operate in the light of all the information they can get, and they need to have this information organized and analyzed in such a way as to give the maximum amount of illumination. But they also need the guidance of that practical wisdom which is born only of experience. What the Federal Reserve Banks need most, therefore, is not more power or less power, or doctrinaire formulations of what their policy ought to be, but merely an opportunity to develop a sound tradition, and to establish it firmly.

A sound tradition would not be the kind of thing which could be formulated in a set of rules and maxims. It would be the kind of tradition which leaves room for growth, for adjustment to constantly changing conditions, and for occasional experimenting. An essential ingredient in such a tradition is the habit of basing policies upon a balanced view of a total situation, and not merely upon features which happen to obtrude themselves in a particular way or at a particular time. Another ingredient of a sound tradition, of course, is that neither political pressure nor pressure from business interests be permitted to affect the policies of the system.

Substantial progress has been made toward building up just such a tradition within the Federal Reserve System. It is more important that such a tradition

should be established than it is that authority within the system should be distributed in one way rather than in another. I suspect, however, that the individual Federal Reserve Banks rather than the Federal Reserve Board are likely to become the repositories and guardians of the accumulated experience and practical wisdom of the system. They are likely, year in and year out, to have the stronger personnel. They are a step further removed from political influence and political pressure. They are directly responsible to the business community as well as to the Government. Most of all, they are in direct and constant contact with the market. They have to make decisions, and they have to act. They learn by doing, and not merely by observation and analysis. The continuity of their tasks and problems makes for the building of an institutional spirit and an institutional tradition.

Nothing which the Federal Reserve Board has to do serves of itself to give its members that thorough education in the difficult and complicated problems of central banking which the responsible officials of the different Federal Reserve Banks get from the tasks that are assigned to them. Individual members of the board may take their responsibilities seriously, or they may not. But there is no way in which the officials of the different banks can escape their responsibilities. Moreover, there are likely to be recurring differences of opinion among the members of the board. Changes in its personnel are likely to be followed by changes in its policies.

In one respect the Federal Reserve Board has an advantageous position. Because it is national, and not regional, because it is removed a little from the immediate problems of the different regional money markets, it may sometimes be able, if it has the wisdom, to take a

broader and longer view. As a coördinator of policies, as spokesman for other interests than those which may come into the immediate view of the various Federal Reserve Banks, as a court of last resort in matters of proposed changes in banking policies, as a buffer between the banks and political influences, the Federal Reserve Board has important functions. It would lose its position of vantage and in the long run it would lessen its own influence if it were to attempt to participate more directly in the actual administration of the system or in the formulating of immediate policies.

II. THE POSITION OF THE NEW YORK BANK

Those who planned the Federal Reserve System confidently expected that the system would operate so as to do away with the "artificial" concentration of banking power in New York. The New York district was purposely made small, so as to include only the up-State banks, which, it was thought, would "naturally" keep their accounts in New York banks rather than in the banks of other cities. The Federal Reserve Bank of New York was to be merely one of a number of regional banks, alike in respect of structure and functions, even if not in respect of size. Each bank was to serve its own region, while the Federal Reserve Board was to supervise and coördinate the operations of the system.

These expectations were bound to be disappointed, for they were based upon a mistaken analysis of the situation that existed in the United States before the Federal Reserve System was installed, and upon a failure to take into account some very important aspects of the operations of central banks in other countries. In the diagnosis of the domestic situation altogether too much importance was attached to the peculiar require-

ments of the National Banking Law as explaining the concentration of banking power in New York. Dr. H. P. Willis, who at the time was expert to the House Banking and Currency Committee, explained in 1913 in a memorandum prepared for President Wilson that "the massing of funds in New York and other financial centers, of which so much has been said in recent years, is largely due to the present reserve requirements of national banks," and that the presumed effect of the new plan would be "to end the placing of reserves with Central Reserve city banks for use in stock market operations, to keep reserves in some measure at home, and to require speculators to get the funds they need in their operations either by directly borrowing them from persons who hold them and want to lend the cash for that purpose, or else by borrowing from the banks in the places where the operations are to be carried on." [1]

Now the massing of funds in New York and the use of part of them in financing Stock Exchange speculation was no new thing. Early Comptrollers of the Currency had referred to these phenomena disapprovingly in their annual reports. But their history goes back beyond the time when the National Banking System was established. Long before that time call loans on Stock Exchange collateral had come to be an important feature of the New York money market, and out-of-town banks had come to keep large balances on deposit in New York and to regard such deposits as supplementary reserves. The National Banking Laws, in permitting New York balances to be counted as reserves by interior banks, merely recognized and regulated existing practice.

The panic of 1857 brought all of the bad features of the system into clear relief, just as another great crisis

[1] H. P. Willis, *The Federal Reserve System*, pp. 174, 175.

did a half-century later. The discussions which followed the panic of 1857 were strangely like those which were prompted by the crisis of 1907. Professor Harry E. Miller, who has inquired into these matters more thoroughly than anybody else, says: "The necessity of maintaining surplus reserves in the New York depository banks, of discontinuing the practice of paying interest on deposits, and of eliminating the use of call loans as a secondary reserve were matters of commonplace knowledge after 1857." [1]

If the setting-up of the old National Bank Reserve System was not responsible for the concentration of banking funds in New York, neither did the establishment of the Federal Reserve System put an end to it. At the present writing the Federal Reserve Bank of New York holds about thirty-five per cent of the total reserves of the Federal Reserve System, thirty-two per cent of its total resources, and forty per cent of its total deposits. On June 30, 1926, the net bankers' deposits (amounts due to other banks minus amounts due from other banks) in the national banks of New York amounted to $909,000,000, as contrasted with an aggregate of $892,000,000 in the national banks of the other sixty-three reserve cities of the country. The total amount held by the national banks of the eleven cities, aside from New York, in which there are Federal Reserve Banks, was $565,000,000; and fifty-five per cent of this total was held by the national banks of Chicago and Philadelphia.

These figures, after all, merely show what every one knows — namely, that New York still retains its unique position as the center in which the surplus banking funds of the country are concentrated. They fail, however, to

[1] *Quarterly Journal of Economics*, February, 1924, p. 329.

give an adequate notion of the amount of the funds thus concentrated in New York. The figures are for national banks only, and out-of-town banks have large accounts in some of the other New York banks. Nor do these figures include the funds which out-of-town banks place directly in the New York market through their correspondents. Then there are the large balances which corporations whose industrial operations are outside of New York keep in New York banks or employ in other ways in the New York market. There is no way of estimating with any precision the total amount of outside funds for which the New York market is currently responsible. But it must be several billions of dollars.

The current figures for the amount of brokers' loans outstanding merely emphasize what had long been obvious, namely, that the new reserve system has not been an effective obstacle to the use of banking funds in financing Stock Exchange speculation. The Federal Reserve Banks, within the limits of their resources, can impound whatever part of the country's stock of gold and currency they think best, and can set limits beyond which credit cannot expand. But, short of exerting a remorseless and destructive pressure on business enterprise, they cannot keep the supply of credit always taut. The law requires that their member banks shall keep certain minimum balances on deposit with them, and prescribes that for such banks these deposits, and these deposits alone, shall count as legal reserves.

But there is no way in which the Federal Reserve Banks can compel even their member banks to deposit with them any idle cash funds that they may happen to have in excess of the reserves required by law. Nothing that the Federal Reserve Banks can do will prevent such funds from flowing to New York or to any other market

where the demand for loans is elastic enough to absorb them. Under any conditions short of extreme strain, such surpluses are bound to appear at various points in the banking system, and there is no practicable way of impounding all of them so as to keep them out of the loan market.

Nor can the Federal Reserve Banks do much to divert the funds available in the market to one use rather than another. The restrictions upon the types of paper which the Federal Reserve Banks will lend upon or rediscount have no effect except so far as they lead commercial banks to discriminate against other types of paper. The money market is not constructed with water-tight compartments. Funds supplied at any one point increase the available supply in every corner of the market. There is no lack of channels of communication.

It should not be inferred, however, that the creation of the Federal Reserve System has had no important effects upon the financing of Stock Exchange speculation. For one thing, call-loan rates are steadier and their range of probable variation is smaller. Open market operations and rediscounting have served to make the supply of funds both more elastic from day to day and steadier over long periods of time. For another thing — and here we come to a point which may have very great importance — it is possible that the existence of the Federal Reserve System has altered the old competitive relation between the financing of speculation and investment and the financing of trade. The point requires explanation.

Before the war, in periods of depression, bankers' deposits, and with them the country's idle cash, accumulated in New York. Except in prolonged periods of depression, such as followed the crisis of 1893, low interest

rates generally led to a revival of long-term financing as well as (commonly a little later) to activity on the Stock Exchange. Some of the funds put at the disposal of borrowers in New York were gradually paid out in different parts of the country (as, for example, in construction expenses) and probably this was often one of the factors that helped business to get on its feet.

With the recovery of industry and trade, funds were withdrawn from the New York market, and the demand for the increased amounts of money that were needed in hand-to-hand circulation also impinged upon New York. The New York banks kept on increasing their loans and discounts as long as they could, but sooner or later losses of deposits and of cash forced them to contract, and this compelled contraction in the securities market.

Before the war it was literally true that a cyclical expansion of production and trade was at the expense of speculation and long-term financing, and it was equally true that business depression created conditions which made an expansion of speculation and of long-term financing possible. This process of cyclical give-and-take between the New York money market and the interior was reflected in the relation between the timing of the expansion and contraction of speculation and the expansion and contraction of general business activity. It probably accounts in part for the way in which the movement of speculation appeared to forecast the movement of business.

The situation which has obtained since 1924, with Stock Exchange speculation and business activities both maintained over a long period at high levels, and requiring, despite falling commodity prices, an enormous expansion of bank credit, had no close parallel before the war. A number of abnormal factors, including con-

tinued gold imports, have played a part in this situation, so that it is impossible to say with certainty just how far the existence of a new and more elastic source of supply of currency and credit has altered the old cyclical relations between Stock Exchange speculation and general business activity. But with the surplus reserves which the Federal Reserve Banks will normally hold, with their ability to supply currency without dipping directly into their reserves, with the power which they have to prevent business activity from expanding too rapidly, it is perfectly clear that those relations have lost much of their former rigidity.

Another important function of the New York market remains to be discussed. It has been and remains our only really important point of contact between the domestic and the international money markets. The men who planned the Federal Reserve System seem to have given little consideration to the fact that the banking reserves of the world are like an international pool, from which the banks of any country may draw. This fact, more than anything else, explains the circumstance that under the old reserve system American banks, whether in New York or elsewhere, rarely held any considerable amount of surplus reserves, except in periods of extreme depression. The accumulation of a surplus in New York, with the resulting fall of interest rates, created a condition favorable not only to the undertaking of certain kinds of financing at home, but also to the reduction of our net international indebtedness, and thus to exports of gold. In the same way the expansion of American business, accompanied by a flow of money from New York to the interior and by advancing interest rates, led not only to the withdrawal of funds from the stock market, but also to an increase of our borrowings

abroad, and thus to gold imports. New York, one might say, was the channel through which variations in the country's monetary demand for gold were transmitted to the international market.

The mechanism of international borrowing and payment imposes upon central banks some of their most important responsibilities and endows them with some of their most effective powers. A central bank located at a strategic point can operate so as to accelerate or retard the international movement of credit and of gold. This has long been recognized in the practice of the Bank of England, whose operations have been guided more by the state of the foreign exchanges than by any other single criterion. The central banks of other countries have generally followed much the same course of action and (allowing for the differences between debtor and creditor markets) have relied upon much the same mechanism. Relatively small differences in the discount rates of the great central money markets of the world, it has been found, are often enough to turn the current of international borrowing in one direction or another, and to start gold movements or to stop them.

Every money market is sensitive in greater or less degree to the operations of central banks in other money markets. In normal times, however, a central bank can take care of the interests of its own national money market, while operating independently of the central banks of other countries. But even then its policies have to be adapted to a situation which is controlled in part by what other central banks are doing. Under exceptional conditions, as when the position of a newly stabilized currency is not yet secure against the pressure of an adverse movement of the foreign exchanges, a central bank may not be able to control the situation by

acting independently, except perhaps by resorting to measures so drastic as to create exceedingly difficult problems.

In the face of just such difficult and delicate situations some of the great central banks of the world have managed during the past few years to achieve a degree of coöperation, or at least of mutual understanding, which, so far as I know, is altogether new in the history of central banking, and which may prove to be the most important forward step in the rational control of credit fluctuations that has been taken since the powers and responsibilities of central banks first began to be recognized. Such coöperation has naturally taken the form of coöperation on the part of the central banks of New York, London, Berlin, and other national financial capitals, rather than on the part of Governments. The central money markets of the world are the only points at which the flow of international credit and international payments can be effectively controlled.

It is in these great markets that the world's demand for and supply of loanable funds meet and are adjusted. This in turn is one of the reasons why it is to these markets that the surplus funds of each country flow and to which each country looks for funds in excess of what can be supplied out of local accumulations. It is in these markets, therefore, that central banks can operate most effectively. The New York Bank has come to have a position of primacy in the Federal Reserve System merely because it has not tried to shun the responsibilities which are naturally assigned to it by the structure of the national money market and by the relations between that market and the world market. Doubtless this involves a departure from the "intent" of the Federal Reserve Act. But such departure was inevitable.

It does not follow, however, that the Federal Reserve Act ought to be amended so as to recognize the character of the situation, as, for example, by putting the New York Bank more directly under government control, or by making the eleven other banks branches of the New York Reserve Bank. The present situation, which is wholly consistent with the law, even if not contemplated by those who made the law, has its advantages.

The New York Bank is merely first among equals. As things are now it is allowed to find its own way of dealing with some (not all) of the different types of problems which confront it, so that it can gradually build up a sound banking tradition upon which the other banks of the country and the business community in general can rely when they are making their own plans. The range of its discretion, however, is limited not only by the control exercised by the Federal Reserve Board, but also by the practical necessity of keeping its policies in harmony with the policies of the other Federal Reserve Banks.

On economic as well as on political grounds it is well that we have a system of regional reserve banks instead of one great central bank with regional branches. The different regions have their own interests and their own problems, and neither the seasonal nor the cyclical variations of the demand for money are alike in all of them. There is room for some degree of regional autonomy in respect of banking policies. Such policies need not always be uniform in order to be elements in a harmonious national system. Sooner or later we shall learn how best to differentiate between the functions and operations of the New York Bank and the eleven other regional banks.

Nothing else that has been done since the Federal Reserve System was established departs so far from the

"intent" of the law as the pooling of the open-market operations of the reserve banks, the centralizing of these operations in New York, and their supervision by a committee of officers of the reserve banks in conjunction with the Federal Reserve Board. Open-market operations are thus undertaken by the system as a whole, and the system operates in the New York market.

The effective powers of the eleven other regional banks are less than those of the New York Bank, merely because they are located in money markets which are smaller, less elastic, and less directly a part of the international money market than the New York market is. Just how far the system as a whole should be further "unified," with the New York market as the center of the system's most important operations, and how far regional autonomy should be preserved, is an important and difficult question. It would be better, I should say, to insist upon the separate regional responsibilities of the different reserve banks, even if this means frankly conceding larger responsibilities and larger powers to the New York Bank than to any of the other banks, than to run the danger of "unifying" the system into a cumbrous and slow-moving substitute for a single central bank, with divided authority and divided responsibilities.

VI

THE CONCENTRATION OF WEALTH AND ITS MEANING[1]

During the last one hundred and fifty years the production of wealth has grown faster than population has grown. New productive methods, new kinds of wealth, new forms of business organization, have followed one upon another in rapid succession. There are few varieties of income which can keep pace, year after year, with this process of restless change. In a dynamic society more is produced than has to be imputed, in the form of income, to the efforts of the rank and file of the productive army. Economic progress yields a disposable surplus. The shares in this surplus are the principal stakes in the game of business enterprise. Here are the roots of some of the most striking phenomena of the concentration of wealth, and here are the widest disparities between service and rewards. There are prizes here for efficiency, for alertness, for foresightedness, and for luck and for unscrupulousness as well. Here are the chief sources of the gains of monopoly, of financial manipulations, and of strategic competitive advantages. Here are found most of those capitalized income-yielding opportunities which we know under such names as franchise values, corporate excess, and good will. The fruits of progress are not apportioned at first among all the coöperating producers, nor do they go, in major part, to the pioneers of science and industry who have

[1] Excerpts from a paper, "Do the Statistics of the Concentration of Wealth in the United States Mean What They are Commonly Assumed to Mean?" *American Economic Review*, vol. VII, no. 1, Supplement (March, 1917).]

made the largest effective contributions to the knowledge that makes progress possible. They go to those who actively and successfully contend for them.

But out of it there come increased demands for labor, for savings, and for productive agents of all kinds. Unless the supply of some productive agent is increasing with undue rapidity nothing can prevent it, as the volume of output grows, from commanding a higher price in the market. Monopoly may crumble and other business advantages may be snatched away by competition, but the forces working toward the diffusion of the product operate relentlessly and surely. Every bit of ground gained by the rank and file is tenaciously held, and becomes a starting point for yet further progress.

This impressionistic picture is in essential harmony with the doctrines of most of the schools of economic theory. Its relation to the interpretation of statistics of the distribution of incomes and of property must be apparent. Such statistics give a cross-section view of things; they speak for a given year or a given moment of time. But the concentration of wealth that exists at any one time in a progressive society is in part an outcome of the circumstance that the productive army does not advance in even formation in its attacks upon the new sources of wealth. There is an advance guard and there are laggards and stragglers. No one knows just how far the concentration of wealth would be reduced, just how far the forces tending toward diffusion would operate effectively, if economic progress should cease, if our dynamic society should crystallize into a static state. That distribution would be much evener is certain. The concentration of wealth is in part one of the unwholesome fruits of progress. If concentration and diffusion are processes going on simultaneously,

statistics show us merely the extent to which, at any one time, the forces making for concentration are ahead of the forces making for diffusion.

It may be objected that this is to no purpose, that present inequalities in the distribution of wealth and the social problems growing out of them are none the less real. It is small comfort to those who are getting less than their fair share of the present product to know that they or their successors are likely sooner or later to get more, particularly if along with this assurance there goes the probability that new inequalities are to be built up as fast as the old ones are leveled down.

Now it is true that to impute these inqualities to one cause rather than to another does not lessen their gravity. But the right diagnosis of a disease has an important bearing upon the choice of remedies. And so the right interpretation of the statistics of the concentration of wealth must strengthen the convictions of those of us who are opposed to an arbitrary leveling down of fortunes or to a revolutionary change in the general structure of our economic life. It gives us confidence that we are on the right road in trying to control monopoly, to restrict the abuses of corporation finance, to do away with unfair privilege and advantages, to establish and enforce fairer standards of competition, to strengthen the bargaining power of the weaker, to reduce the extent to which accrued inequalities are transmitted from one generation to another by inheritance, and, by controlling immigration, to check the tendency of the process of diffusion to become a process of dilution. It may well be that this road is longer and has more turnings than we now imagine. But there is nothing in what we now know about the distribution of wealth to suggest that this road — the road toward

equality of competitive opportunity — is not the right road.

I pass to a more technical question. What is the precise meaning of the concentration of wealth? By what standards shall we measure it? Statisticians have generally used "concentration of wealth" and "inequality in the distribution of wealth" as loosely interchangeable terms. Now any departure from perfect equality in the distribution of wealth means *inequality*. But is *concentration* to be defined so broadly as this? Wealth might be distributed unequally, without there being any amassing or concentration of any relatively large part of it in the hands of any one group or portion of society. Concentration means, then, a particular kind of inequality in distribution. And moreover, while any degree of centralization must be deemed concentration, yet the social problem of the "concentration of wealth" is the problem of its undue or *excessive* concentration. But we have no definite standard of what constitutes justifiable, permissible, or normal concentration. And so the statistics are made to indicate merely the gross departure from a condition of absolute equality in distribution. One has to be on one's guard, therefore, against imputing to them a significance which possibly they may not have.

There ought to be agreement, it would seem, about what constitutes *equality* in the distribution of wealth. Pareto's interpretation of the meaning of his own well-known index of inequality in the distribution of incomes is based on the expressed assumption that when the number of persons with incomes of less than a given size increases as compared with the number of persons with higher incomes, the inequality of income distribution

diminishes. I do not quarrel, as some have, with this definition of relative equality and relative inequality in distribution. Given a definite income range, with small incomes more numerous than large incomes, Pareto's hypothesis is, in general, consistent with the common notion of the meaning of equality in distribution. But Pareto made a curious slip in interpreting the relation of his index to his definition of inequality.[1] His index does not increase, as he supposed, with what he deemed inequality in the distribution of incomes, but decreases. It can be used, as Benini and Bresciani have used it, in this inverted fashion. In its direct form it might conceivably be taken as a rough measure of the evenness with which income receivers are distributed through the income range. But it gives no simple and definite standard of comparison. The index becomes unity, for example, when the relative numbers of the persons receiving incomes of stated sizes are inversely proportional to the squares of their incomes. How far, in general, the notion of an even distribution of income receivers throughout an income range differs from the common notion of equality in income distribution may be inferred from the fact that, with an absolutely even distribution of income receivers over *any* income range, the richest fourth of the population would get seven sixteenths of the total income, while the poorest fourth would get but one sixteenth.

For the most part, however, equality of distribution is interpreted literally; that is, it is taken to mean absolute uniformity in the distribution of income. Thus, when a statistician throws his estimates into the familiar form that assigns a certain (large) proportion of the aggregate income to a certain (small) proportion

[1] *Cours d'Économie Politique*, vol. II, p. 320, note.

of families, the comparison inevitably implied is with a state of things in which 50 per cent of the families get exactly 50 per cent of the aggregate income and 10 per cent of the families get 10 per cent of the income. And so with Dr. Lorenz's graphic device for representing the way in which such proportions depart from the line of absolutely equal distribution. So, too, with the index of concentration which Professor Corrado Gini has suggested as a substitute for Pareto's, but which increases when Pareto's decreases, and which becomes unity when one income receiver gets just as much income as another. So also with Gini's other index, which takes into account the sum of the differences between each income and every other income. This index, it is interesting to note, may be interpreted as a summary arithmetical expression of the degree of concentration denoted by the Lorenz graph. So, finally, with all of those measures, such as the average deviation, probable error, and standard deviation, which indicate in a general way the extent to which incomes as a whole differ from the average income. The significant thing is that all of these ways of expressing the degree of inequality in the distribution of wealth use as a standard or reference of comparison an absolutely equal and uniform distribution.

Some or all of these measures are useful in comparing the distribution of wealth in different countries or at different periods. But none of them is of much help in forming a judgment with respect to the degree of undue or excessive concentration that may exist. The degree of departure from absolute equality, however measured or stated, must itself be referred, if not explicitly, then in some vague way, to a standard of normal or justifiable concentration. A dead level of uniformity

is neither practicable nor desirable as an ideal of distributive justice.

A concrete example may give point to this consideration. Suppose that incomes in an imaginary society were distributed symmetrically around the modal or most common income, in the form of a normal frequency distribution. This might represent either one of two things: (1) a normal distribution of ability and a perfect proportioning of income to ability; (2) a random or chance distribution of incomes, under the influence of complex but unbiased forces. This second condition would be consistent with the existence of real equality of opportunity, broadly understood, coupled with the presence of a myriad of small circumstances that might deflect an income receiver towards a lower or a higher portion of the income range. Now suppose that the average family income is $1500 and that half of the families get incomes that are within $200 of this average. Under such conditions the richer half of the families would get 58 per cent of the aggregate income and the poorer half would get 42 per cent. Increase the dispersion of the distribution somewhat, so that half of the incomes are between $1000 and $2000. Then 70 per cent of the aggregate income would go to the richer half of the population, and 30 per cent to the poorer half. Increase the limits between which half of the incomes fall to $800 and $2200, and the portion of the aggregate income assigned to the richer half of the population becomes 78 per cent, leaving 22 per cent for the poorer half.

Dr. King's recent estimates assign about 27 per cent of the aggregate income to the poorer half of the families and 73 per cent to the richer half. This is a slightly *smaller* degree of concentration than would be given by

a normal frequency distribution with half of the incomes falling between $900 and $2100. This suggests that no single or general statement of the degree of concentration, taken by itself, is adequate.

Instead of tabulating the statistics in the misleading form of the proportions of aggregate income or property in the hands of stated proportions of the population, it is better to use a simple frequency distribution, showing the relative numbers of income receivers or property owners in the different income or property classes. Such frequency distributions can be adequately described and compared, one with another, and with various ideal schemes of distribution by means of the constants statisticians use to measure spread, skewness, and curvature. Such a handling of income statistics serves to focus attention upon the really important things, which are the upper and lower limits of the income scale and the manner in which income receivers are distributed between these limits. The *amount* of concentration does not matter so much as does the particular *form* of the income distribution underlying the concentration. An identical general degree of concentration may result from a fairly good and a very bad distribution of incomes.

The worst thing in the present situation is undoubtedly the extreme skewness of the income frequency curve. The mode — the most common income magnitude — is very close to the lower limit of the distribution. Then the income curve descends rapidly as the higher incomes classes are brought under review, reaching a condition of relative attenuation at incomes of only a few thousand dollars, but stretching on for an absurdly great distance before the maximum incomes are reached. The problem of poverty and the problem of great fortunes are

the problems of the lower and upper limits of this income curve. But the problem of great fortunes is only part of the larger problem of the general skewness of the curve, the problem, that is, of the extremely small average differences in the incomes of persons in the lower part of the income range and the unduly rapid increase of these average differences as the view is shifted to successively higher income groups. Put concretely, that 10 per cent of the families in the country get possibly three fifths or two thirds of the aggregate income ceases to appear principally as a problem of large fortunes, when it is realized that to include the richer 10 per cent of the families, one has to be down to somewhere between the $1200 and $1800 income levels.[1] The most serious aspect of the distribution of property and incomes in this and other countries is not the presence of a larger or smaller degree of "concentration," but the general distortion of the whole income scheme, reflecting as it undoubtedly does the presence of a high degree of inequality in the distribution of opportunity.

I propose now to discuss two minor points which have an important bearing upon the degree of inequality reported by the statistics of the distribution of wealth and incomes.

In the first place, what disposition shall be made in income statistics of the appreciation and depreciation of capital values? I do not refer to such appreciation and depreciation as come from improvements and better-

[1 Estimates of the distribution of incomes in the United States in 1918, made by the staff of the National Bureau of Economic Research, Inc., indicate that the most prosperous 10 per cent of the income receivers had less than 35 per cent of the total income, and that an income receiver was in this prosperous 10 per cent if his income was as much as $2400. See *Income in the United States, 1909–1919*, vol. 1, p. 147.]

ments on the one hand and wear and tear on the other, but to the changing money value of unchanged things. These changing capital values, it is true, are in large measure a result of changes in the probable *future* income-yielding power of the things valued. But it does not follow that to count appreciation as income involves double counting. If savings are to be counted as income, and they must be, for the purposes of income statistics, appreciation in the money value of income-yielding goods or rights, being closely akin to savings in a number of important respects, must also be counted as income. It is no part of the annual product, as usually defined, but it is part of the annual *distribuendum*. It enters into the annual changes in the relative economic positions of the different members of society. Any one income receiver can convert his share, if he wishes, into money income and then into real income. It is unnecessary to distinguish between such appreciation as merely records the shrinkage in the purchasing power of the dollar and such appreciation as comes in other ways. So long as appreciation, of whatever sort, is distributed in other manner than in proportion to the other elements in current incomes, it has to be taken into account in income statistics, or at least in the interpretation of the significance of such statistics.

The census estimate of national wealth for 1912 exceeds the similar estimate for 1900 by about a hundred billion dollars. Some of this — I doubt that it is more than a third — represents an increase in the national inventory measured in terms of physical units, together with the increase in the unit value of short-lived equipment and products — the current assets of the nation. The balance of the gain consists largely of net appreciation in the value of land and other durable forms of

wealth. And there is an enormous amount in franchise values, monopoly rights, and good will and other vendible competitive advantages that does not enter into the statement. Altogether, it is likely that the recent average annual addition to the national *distribuendum* on account of net appreciation has amounted to as much as five and possibly as much as seven billion dollars, making a difference of perhaps 20 or 25 per cent in the national income statement.

One of the defects of British income statistics is that they do not cover gains from increases in capital values, even when these gains are realized in actual sales, except where speculation is carried on as a business. Now the effect of the inclusion of these items upon the apparent degree of inequality in the distribution of incomes will depend very largely upon whether they are counted as income only when realized by actual sale or set down yearly as accruals. Take the case of a farmer who bought a hundred-acre farm ten years ago. Since then it has been increasing in money value, as many farm lands have, at the rate of $2 per acre per year. He now sells it with a gain of $2000. If all of this profit is set down as income for this year, it lifts him, for the once, into the higher income groups and increases the general skewness of the distribution of income for the year. But if accruals had been counted at the rate of $200 a year for ten years the increase might not have carried him up above the average income for the country. In fact, the skewness of income distribution might very possibly have been reduced. If there were a million other farmers similarly conditioned, there would be the choice between counting a million moderate gains or an average of perhaps a hundred thousand large gains annually. And so with

those enormous increases in the values of business equities which have been a very important source of large fortunes, the method of accruals will increase the skewness of the income curve very much less than will the method of realized sales. For the purposes of income statistics there can be little doubt that the method of accruals should be preferred. To wait until the farmer's increased land values are realized in sales exaggerates the farmer's real position in the income scale in one year out of the ten and understates it the other nine years.

My other point relates to the statistics of the distribution of property. The most important facts that we have, and the facts most frequently quoted, have been gleaned from probate records. I have little confidence in the inferences that have been drawn from these figures. The principal difficulty is the difference between the age distribution of decedents and the age distribution of the living population. In England, for example, while the average age at death of that part of the population which lives to be at least twenty years old is over sixty years, the average age of living persons over twenty is about forty. Now as men grow older they often grow richer, and this fact has been shown by Mr. Bernard Mallett to have a very important bearing upon the proportion which the value of the estates annually passing from one generation to another makes of the aggregate capital value of the national wealth. For example, in the fiscal year 1913–14, the average value of the taxable estates left by persons dying in that year was £4150. But those dying at between twenty-five and thirty-five years of age left estates of an average value of only £1600. The average

values of the estates which had been accumulated by persons dying between the ages of sixty-five and seventy-five was nearly £4200, while persons who lived to be more than seventy-five left estates with an average value of over £6000. In the United States, where fewer persons start with the initial advantage of inherited wealth, these differences are likely to be considerably greater. Now the danger is not that the probate records will give us an exaggerated idea of the aggregate money value of property holdings, for we have other and better sources of information on this point. But they give an impression of a higher degree of inequality in the distribution of property than really exists. Not only do many men grow richer as they grow older, but some men grow rich faster than others, while some men, especially among those with little property, grow poorer as they grow older. The inequality of possessions among persons at the end of life is very much greater than among the living population. Statistics of the sizes of estates admitted to probate are nearly worthless unless they are accompanied by statistics of the ages of the decedents. And with the probate statistics thrown out of court, there remains virtually nothing that gives us any adequate notion of the distribution of accumulated wealth in the United States.

VII

PERSONAL AND IMPERSONAL TAXATION[1]

All taxes are personal in the sense that in their final incidence they are paid out of personal incomes; all taxes (except the poll tax) are impersonal in the sense that they are apportioned on the basis of some impersonal magnitude, such as property, income, product, import, or transfer. But there is a real difference between what we may properly call personal and impersonal taxation, that is, between the taxation of persons and the taxation of things.

The distinction at its broadest is between taxation that falls evenly upon a given class of things, without reference to their ownership, and taxation distributed uniformly over a given class of persons. Thus indirect taxes, whatever their ultimate incidence, are from this point of view always and everywhere impersonal taxes. They are imposed upon various classes of things — goods or transactions — without primary reference to the taxpayer, to his relation as resident or non-resident to the Government imposing the tax, or to his taxpaying ability. So with import duties (in their revenue aspect), with excise taxes, with stock-transfer taxes, with stamp taxes of various sorts, and with a host of others.

But while all indirect taxes are of necessity impersonal taxes, seeking the taxable thing or the taxable act, and disregarding the status of the individual taxpayer, it does not follow that all direct taxes are personal taxes. Income taxes, proportional or progressive, falling alike

[1] Reprinted, slightly abridged, from the *Proceedings of the Ninth Annual Conference of the National Tax Association* (1915).]

on different classes of income-receivers, are personal taxes. The general property tax is a personal tax in its general purpose and its traditions, although not in all the details of its operation. And so with the inheritance tax and the habitation tax. But a tax upon one specific kind of property, such as a tax upon automobiles or dogs or land, or even upon real estate in general, is not, taken by itself, a personal tax. Such a tax may, indeed, if supplemented by taxes upon other evidences of tax-paying ability, be part of a general system of personal taxation. Isolated, however, it is an impersonal tax.

Personal taxation rests more or less solidly upon the ability-to-pay principle of taxation; impersonal taxation finds a slender prop in the benefit principle, or expresses national commercial policy, or is to be considered as sumptuary legislation, or frankly commands support on grounds of fiscal expediency and administrative economy, or, as in the case of the single tax on land, is based upon an alleged public equity in some particular form of property. This is not to say that all impersonal taxation is bad. In particular circumstances one or another of these various principles may have both validity and weight. There is small doubt but that we shall continue to have impersonal taxes, and perfectly defensible ones, too, and very likely in increasing variety. The pressing question, as I see it, is to what *extent* these two contrasting principles — personal or impersonal taxation — shall be recognized in our tax systems. More particularly, the question is whether the principles of personal or of impersonal taxation shall be accorded supremacy in the work, now going on, of remodeling, pruning, and supplementing the general property tax.

The general property tax, as I have said, is in principle

a scheme of personal taxation. In its origins property was taken as the measure of the taxpaying ability of the property-owning citizen. It was possible to keep practice fairly well in line with principle so long as nearly all taxable property was found in concrete and tangible forms — lands, building, and goods — and so long, moreover, as such property was located, for the most part, in the taxing districts in which its owners lived. This primitive situation has been changed. Wealth still is found in concrete and tangible forms, multiplied both in quantity and variety. But the old intimate and direct relations between such wealth and its owners is rapidly passing. In place of simple and direct forms of property and possession we have a complex subdivision of the equities in industrial wealth, represented by a mass of securities and other evidences of rights of one sort or another. And these securities, these evidences of rights in the uses and benefits of wealth, are themselves property. So we find that the distribution of persons, of taxpaying citizens, among different taxing districts is a very different thing from the distribution of concrete and tangible forms of wealth, while both are different from the distribution of the securities that represent the equities in this wealth.

It is this continuing process of the subdivision and dispersion of property rights that has brought us face to face with the problem of personal or impersonal taxation. In this separation of men and their possessions shall taxes follow the man or the property? The easiest and probably the worst way out of the difficulty was merely to ignore the problem, and to refuse to lift the burden of taxation from either the person or the property. At any rate, that is what happened. And so we have the problem of double or multiple taxation by com-

peting tax jurisdictions, the problems connected with the taxation of securities and credits, and, as I think I may add, many of the problems of corporation taxation. These are the problems with which we are wrestling today; and hardly a proposal has been made toward solving them which does not involve some definite attitude toward the general question of personal or impersonal taxation. At one extreme are those who would frankly throw aside all attempts at personal taxation and would depend for state and local revenues upon taxes on tangible forms of wealth, especially real estate, supplemented, perhaps, by a variety of indirect and special taxes. Back of many such proposals lies the shadow of the single tax, which involves, of course, the frank abandonment of any personal basis of taxation. At the other extreme are those who would retain personal taxation at all costs; who would enforce the general property tax upon all kinds of property rights, despite the double taxation inevitably involved, or who would supplement the tax upon tangible property by special taxes upon intangible property or upon incomes.

The advocate of impersonal taxation is not without weighty arguments. Simplicity, convenience, efficiency in raising revenue, and similar points are things which he is prone to stress. Then, particularly if he is from a region where much outside capital is invested, he will very likely mention the right of the state or locality to tax wealth which is entrusted to the protection of its laws. And if he is a single-taxer he will urge that there is a public equity in land values, and will talk about the repressive effect of taxes on other kinds of wealth.

The advocate of personal taxation, in his turn, will have no difficulty in making out a strong case. Personal taxation is an expression of the now dominant

ability-to-pay or faculty theory of taxation. It is in line with the origins and traditions of voluntary taxation among English-speaking peoples. There is a wholesome social ideal embodied in the notion that each member of the community should contribute to the public purse in accordance with his means. Personal taxation serves as a check upon the abuse of the taxing power and stands in the way of that extravagance in public expenditures which impersonal taxation is apt to encourage. It recognizes the right of the community to public contributions from its wealthy citizens, whether they be local property-owners or not.

This second set of considerations seems weightier than those which I have mentioned as supporting the case for impersonal taxation. Yet we must face the fact that anything like a thorough-going application of the principles of personal taxation is impracticable and undesirable.

In the first place, there is the obstinate fact that there is no very close relation between the distribution of personal taxpaying ability among different tax districts and the needs of those same districts for public revenue. Anything like a thoroughgoing system of personal taxation would curtail public revenues in many Western States, in manufacturing towns, and in all places where any large amount of outside capital is invested. Public revenues would be relatively redundant in localities where wealthy taxpayers are congregated. The wealthy suburban community of tax-dodgers owes its existence not only to the underassessment of personal property in such communities, but also to the fact that such a concentration of personal wealth operates automatically to reduce the local tax rate. If all taxes were personal

taxes, assessed at the domicile of the taxpayer, we may be very certain that the tendency toward the building-up of such communities, toward the withdrawal of wealthy men from the civic life and responsibilities of poorer communities, would be strengthened. And surely this is objectionable on other than fiscal grounds. Our tax system ought not to further the segregation of different economic classes.

In the second place, any attempt to put all state and local taxation on a personal basis would encounter the difficulty that the most firmly entrenched part of our present system, the tax on real property, is not a personal tax. Of course in the majority of cases both the owner and the property are found in the same tax district, and in such cases the tax on real property fulfills in large measure the purpose of a personal tax. But exceptions are important enough to disprove the rule. Levied as it is at the situs of the property the tax on real property is in principle a distinctly impersonal tax. When the taxpayer and his property are in different States or even in different communities in one State, the impersonal character of the tax stands out clearly. The citizen of one State or community contributes to the support of another State or community if he happens to own real property in it, and in that measure his tax-paying ability is turned into channels that supply other public needs than those in which he, as a citizen, is vitally and personally interested. And yet this tax, taken as a part, even the important part, of a general system of taxation has almost no enemies, even among the most active defenders of personal taxation. It does not square with the ability-to-pay theory of taxation, and so we are accustomed to try to reconcile our general principles and our practice by grudgingly admitting a

modicum of truth in the benefit theory, or in what Mr. Lyons, in his admirable little book, calls the cost-of-government theory, or by talking, in looser terms, of the right of the State to a revenue from the income from all property within its boundaries, or by holding, with the single-taxers, that there is, after all, a legitimate public or community equity in landed property.

In short, there is no consistent recognition of the principle of personal taxation in our present tax practice, and there are decisive obstacles in the way of reorganizing our tax systems upon a definitely personal basis. An examination of the inheritance-tax situation would have led to precisely similar conclusions. The same may be said for the corporation tax, although its personal and impersonal aspects are somewhat more complicated. Personal taxation, then, is a tradition of the past and an ideal for the future, rather than a dominant factor in the present situation.

The one clearly personal element in the general property tax has been the tax on personal property levied at the domicile of the owner. This is why we have clung to it, despite its general ineffectiveness, and despite the element of double taxation involved in the taxation of intangible personal property. This is why, in some States, we still try to enforce it in its crude form, and why, in other States, we give it up only to make way for some substitute for it which will reach the same taxpayers, such as a low tax on the same kind of property or a tax on incomes derived from such property. Of course other factors are also responsible for the effort to find an efficient substitute for the personal property tax. The continuing need of increased public revenues and the unwillingness of real-estate owners to bear

larger tax burdens than they are now carrying, for example, have had an important influence upon the matter. But unless I am greatly mistaken the most powerful single influence back of these new developments in American taxation has been the widely shared conviction that men of means should not be freed from the necessity of making personal contributions to the expenses of their own local and state governments, even though the income of such men might come from property already taxed in another jurisdiction; from equities in business corporations subject to general and special taxation, or from mortgages on fully taxed real property. Because of the pressing importance of this matter of personal taxation we have virtually relegated the problems of ultimate incidence and of double taxation to a position of secondary importance.

The new taxes which are being introduced in place of the old personal-property tax find their justification in the principle of personal taxation. They are personal, but they are tacked onto a system which, in principle, is impersonal. We tax real property at its situs, we tax some kinds of tangible personal property at situs or domicile, or both, and then we tax at the domicile certain kinds of personal equities in property taxed at situs. This statement holds true of a supplemental income tax, such as Wisconsin's, just as it does of a special tax at a low rate on intangible personal property.

Contrast this system with a possible one in which the two bases of taxation, the personal and the impersonal, should be consistently and logically recognized throughout. There would be recognition of the right of the State or locality to levy taxes upon all property located within its borders, and there would be equal recognition of the right of the State or locality to tax every resident

on the basis of his property, wherever located, or his income, whatever its source. A corporation would be taxed on its tangible property in the State or States in which such property is located. Its bondholders would be taxed (at a lower rate, of course) on the securities in the States in which they reside. Furthermore, stockholders would be similarly taxed on their stock.

Yet further — and here is where the scheme goes beyond present tendencies — the landowner would be taxed on his land in the State in which it is located and on the ownership of it or the income from it in the State in which he lives. If owner and land are in the same taxing district, there should be two taxes imposed, one because the land is located there, and the other because the owner lives there.

We cannot transform our present system of taxation so that it shall have an exclusively personal basis; but we can so extend our present personal taxes, which are now superadded to the tax on tangibles in a few small patches, so that they will constitute a complete and harmonious system, absolutely coterminous with the system of the impersonal taxation of wealth at its situs. Double taxation would be eliminated by the simple device of making double taxation the uniform rule.

I am not sure that such a result could be achieved by extending the principle of the two- or five-mill tax on intangibles so as to cover all similar equities, including the landowner's equity in his own property. The courts might hold that the personal tax on the faculty of the landowner was virtually a tax on the land, and hence collectible only at situs. But the general result could be reached just as well by a state income tax, apportioned in part to localities, and falling upon *all* incomes above a low exemption limit. This should not be

a special or supplementary income tax, but an *inclusive* tax upon all incomes received by residents of the State, even where such incomes are derived from property already taxed in the same or other jurisdiction. The tax on real property should be retained in its present form, and such tangible personalty as is taxed at all should be taxed where located.

This is one way, and a practicable way, to bring order and consistency into our systems of state and local taxation, to rid ourselves of double taxation, whether by one or by competing jurisdictions, to retain the very real advantages that go with the taxation of tangible property at the place where it is located and at the same time to introduce a thoroughgoing system of personal taxation, reaching every taxpaying citizen of the State.

I fear that we are letting the pressure of erroneous popular notions of taxation, bred of our familiarity with the general property tax, and the urgent demand for increased state revenues, force us into paths that are bound to lead to confusion, and perhaps to a repetition of some of the worst phases of our experience with the general property tax. It is altogether inconsistent with the principles of personal taxation to apply the state income tax or the special tax on intangibles to corporations, or to make of the income tax a tax on "business." It will inevitably lead to new and serious difficulties (including multiple taxation and retaliatory legislation) if the State attempts to tax foreign corporations upon the basis of the proportion of the business of such corporations done within the State. If "business" must be given a special habitat for purposes of taxation, apart from the domiciles of its owners and the situs of the property it uses, it should be effected by simple and uni-

form rules, applying to domestic as well as to foreign business undertakings. And the imposition should take the form of a privilege tax or license fee, and should not be disguised as an extension of general taxes upon property or upon persons.

VIII

DEPRECIATION AND RATE CONTROL[1]

Two questions outrank in practical consequence all other problems of procedure in the valuation of the properties of public service companies for purposes of rate control. First, what is a proper rate of return upon the investment? Second, shall the property taken as evidence of the investment be valued for that purpose as though it were new, or shall an allowance be made for the fact that it is in various stages of age, wear, and obsolescence? Under American conditions the difference between original cost and replacement cost as a standard of valuation is not likely to be large,[2] except for real estate holdings, and in the aggregate the difference between the results got by the use of these two methods will usually be less important in its bearing upon the determination of reasonable rates than is the margin of uncertainty with respect to either the proper rate of return or the matter of depreciation. Nor do the minor problems suggested by such phrases as "inter-

[1] Reprinted, with minor changes, from the *Quarterly Journal of Economics*, vol. XXVIII, no. 3 (May, 1914).]

[2] The common practice of averaging prices over a period of five or ten years in determining replacement cost contributes to this end. [The advance of prices since 1914, when this paper was first published, has put a new face upon the problem. Reproduction cost is now generally much larger than original cost, and has lost most of its usefulness as evidence of "the amount of the investment entitled to a return." (Below, p. 145.) In this paper I look at both reproduction cost and original cost as affording merely working methods, not principles, of valuation. I assume throughout that whatever technique it employs, the object of public valuation is to determine the amount prudently invested in providing a public service. I have explained this more fully in a note on "Depreciation and Reproduction Cost," in the *Quarterly Journal of Economics*, vol. XLI, no. 2 (February, 1927).]

est during construction," "going value," "donated property," and the like, usually compare in importance, even in the aggregate, with the two that have been named. The present paper deals with the second of these two major issues of valuation.

In a valuation made for the dual purposes of rate regulation and taxation in Wisconsin the present (depreciated) value of the physical properties of twenty-six electric railways was found to be $7,826,000, or 82 per cent of the "cost of reproduction new" of $9,596,000.[1] For individual companies the corresponding proportion was as low as 57 per cent, and was below 79 per cent in as many cases as above it. And these figures include the value assigned to land and grading, from which no deduction for depreciation was made. The depreciated value of the properties of fifty-four steam railroads was estimated as $196,239,000, or 80 per cent of the cost of reproduction new, which was given as $244,129,000. This percentage varied for individual railroads from 61 to 96, the median being 77.[2] Leaving out of account the items of land and grading, on which no depreciation was reckoned, the depreciated value of the remaining assets of the fifty-four railroads was only 73 per cent of the value new.

The Michigan valuation of the properties of steam railroads in 1900 gave a depreciated value that was 81 per cent of the cost of reproduction.[3] In the Minnesota appraisal of 1907 the corresponding percentage was 88.[4] But when the unusually large list of items in the

[1] *Fifth Biennial Report of the Wisconsin Tax Commission* (1910), Appendix D.

[2] Computed from a table in the *Fourth Biennial Report of the Wisconsin Tax Commission* (1909), p. 128.

[3] *Bureau of the Census Bulletin*, no. 21 (1905), p. 78.

[4] *Twenty-Fourth Annual Report of the Minnesota Railroad and Warehouse Commission* (1908), p. 52.

latter valuation against which no depreciation was charged [1] is eliminated from the reckoning, the proportion is reduced to 76 per cent.

Examples of this sort might be multiplied, but it is better to turn to the general principles underlying the aggregate amount of depreciation of the properties of a large business. Mr. James E. Allison, formerly a member of the St. Louis Public Service Commission, as well as its chief engineer, was the first to appreciate the significance of these general principles.[2]

At any given time in the history of a large business undertaking with extensive and varied properties, properly maintained, some of the items of plant and equipment will be new, or nearly new. And unless the business has been newly undertaken, other items of plant and equipment will have nearly served their term of profitable use. Still others will be found in each of the possible stages of wear intermediate between newness and final replacement. It follows that an efficiently maintained property of this sort which has been operated for a sufficient number of years will normally be about half "worn out" and that, if renewals continue to be promptly made when needed, the property will normally tend to remain in this same average physical condition. This conclusion is subject to a few very simple qualifications. (1) No single item of plant or

[1] Including land, grading, track laying, stores and supplies, contingencies, and interest during construction.

[2] *Should Public Service Properties be Depreciated to obtain Fair Value in Rate or Regulation Cases?* Report to St. Louis Public Service Commission, 1912. Reprinted as Appendix A of the *Report on the United Railways Company*, by the St. Louis Public Service Commission, 1912; and as Appendix D of the *Report on the Southwestern Telegraph and Telephone Company*, 1913. Cf. also Mr. Allison's paper on "Depreciation," *Annals American Academy of Political and Social Science*, vol. LIII, p. 198 (May, 1914).

equipment can be so large and expensive, relative to the other items, that its degree of newness disproportionately affects the average state of wear. (2) The various items of plant and equipment must have been acquired at different times or else (the more common case) must have differing periods of life, so that replacements are distributed with some evenness through successive years. (3) The business must be in a static condition. If it is growing there will be a relatively larger amount of new equipment; if it is declining, some of the equipment may not be replaced as it wears out, and the property as a whole will be in a sub-normal condition.

This last condition — that the aggregate size of plant and equipment is supposed to be neither increasing nor decreasing — is one which rarely holds true under American conditions. Most American railroad and public service corporations increase their investment in fixed capital from time to time.[1] But this fact does not affect the significance of the notion of "normal average depreciation." One might, if one pleased, modify the assumption of a normal depreciation of 50 per cent so as to allow for the effects of any given rate of growth. But all practical ends are served by considering the present and future depreciation and probable replacement of the body of assets which exists at a given time. For this purpose the 50 per cent average is generally appropriate. The other conditions are not usually far out of line with

[1] This is one of the reasons why the depreciated value of railroad properties as found by various commissions is generally considerably more than 50 per cent of the undepreciated value. Other reasons (aside from a possible large margin of error in the work itself) are the inclusion of non-depreciating assets in the valuation, and the fact that the residual or scrap value of the discarded assets has to be taken into account. Thus a steel rail costing $28 per ton and with a scrap value of $12 per ton would when "half worn out" be valued at about 71 per cent of its cost new.

the facts. Indeed, the customary endeavor of every properly conducted business to prevent too large a burden of maintenance or replacement expenses from falling within any one year works potently toward securing a normal average condition of depreciation.

Durable capital goods are said to "depreciate," while less durable ones are "used up" or "consumed." The difference is one of degree only. Buildings, rails, and locomotives are worn out slowly; fuel, oil, and other supplies are used up rapidly; but in the long run the two processes are alike. Just where the line shall be drawn between them for accounting purposes is determined by the fact that a year is the ordinary fiscal unit of time. Assets whose normal length of life is more than a year are "capital assets," and "depreciate"; assets of a more transient sort are used up and replaced within the year, and their cost is counted under the head of operating expenses. If the customary fiscal unit of time were five years, or twenty years, there would be fewer assets whose depreciation would have to be taken into account. Part of what are now capital assets would in a longer fiscal period appear as current assets. But, as it is, we wish to know the profit or loss of a business undertaking during a year, and its general condition at the end of the year. So some method must be devised of allocating the cost of the more durable instruments of production among the successive years of their period of life. "The more permanent form of assets serving for productive use during a period of years should be spread as an expense during the period of use, whether that be five or fifty years." [1]

[1] H. R. Hatfield, *Modern Accounting*, p. 122. J. S. Eaton (*Handbook of Railroad Expenses*, p. 3) says: "So we see that capital expenditure, as

It is not surprising that there is no general agreement among accountants respecting the proper basis of thus cutting up and parceling out the total cost of a durable good. The purpose of accounting is to ascertain and exhibit the true condition of a business, but just what the truth is will depend upon the specific question which we ask the accountant. Do we want to know from the balance sheets whether the productive efficiency of the durable equipment of a business is being maintained unimpaired? Is the most accurate statement possible of the profit or loss of a particular year the desideratum? Are we looking toward the future and asking whether adequate provision is being made for the amortization of the cost of the durable capital assets that will have to be replaced? Or do we have a possible insolvency in mind, and seek to know the market value of either the equipment as a whole or its different parts?

Emphasis upon one query would logically lead to apportioning depreciation among the various years of life of a durable good according to the diminution of its productive efficiency; the proper answer to others would be given by apportioning depreciation according to the requirements of annual contributions to a sinking fund or according to the diminution in the estimated market value of the assets in question. Accounting practice often represents a compromise between two or more

distinguished from expenses, is at last an arbitrary conception. It begins with the idea that certain expenditures have an efficiency that reaches over many earning periods extending indefinitely into the future. But nothing physical would last so long, and its earning power might have even less permanence. To meet this condition we arbitrarily designate certain expenditures whose effect indefinitely outlasts the immediate earning period as 'capital,' and then in the same arbitrary way, through all subsequent vicissitudes, we hold them to their first value, by maintenance, renewal and depreciation charges which are borne by current expenses."

different possible methods of procedure. For example, one of the points sometimes made in favor of the method of dividing the total depreciation of an asset into equal annual parts is that while the market value of many capital assets (machines, perhaps) declines most rapidly in the first few years of use, the technical efficiency of the same assets declines most rapidly in the years just before they are discarded.

These and similar considerations have to be weighed in any thorough analysis of the problem of the proper recognition of the fact of depreciation in private enterprises. They are also properly to be taken into account in discussing the business policies of unregulated public service industries. Although the principle that the charges of such industries must be reasonable is of long standing, this had no bearing upon the depreciation problem until some basis for the determination of "reasonableness" was laid down by commissions and courts.

But for regulated public service undertakings the proper adjustment of depreciation accounts is not merely a matter of business policy nor of the proper adjustment of the interests of the holders of the various classes of equities in the business. It directly affects the issue between the proprietors of the undertaking on the one hand and the public on the other. It does this by operating on, first, the amount of annual net earnings shown by the accounts, and, second, the real or apparent magnitude of the capital investment. This bringing in of the public interest as a matter of primary concern helps to define and narrow the problem and at the same time raises new issues.

There is no better way of approaching the question of the proper treatment of depreciation in the valuation of

the properties of public service companies than by examining in some detail the functions and results of depreciation accounts in public service undertakings. Railroads will serve as the best example, not only because of the magnitude of the interests involved, but also because the accounting rules of state public utility commissions have been patterned in large measure after those of the Interstate Commerce Commission. The principles involved are general.

The Interstate Commerce Commission, acting under the increased powers over the accounts of railroad companies given it by the Hepburn Act of 1906, required that the depreciation of rolling stock be counted as an operating expense. The introduction of this account was in line with a general attempt to draw the line more accurately between "additions and betterments" on the one hand and the mere replacement or maintenance of existing capital on the other. Maintenance accounts, as every one knows, had been the loose ends of railroad bookkeeping. In an earlier period of American railroad finance, inadequate appropriations for maintenance of way and equipment furnished a common method of milking the property and of manipulating the value of securities through the payment of unearned dividends. In later years this practice had become less frequent, and many of the stronger roads had gone to the other extreme of overstating their maintenance costs by counting many sorts of additions and betterments as replacements. It was in part to control this practice that the regulations referred to were instituted by the Commission. In addition to these requirements, elaborate rules were made for the definition and classification of additions and betterments.

There was much complaint that these regulations interfered in an unwarranted way with the companies' rights to determine their own financial policies, and much was said about the advantages of the conservatism expressed in the understatement of net earnings. It was alleged, furthermore, that replacements and renewals fell so evenly in the case of a large railroad that it was unnecessary to provide for them in advance. The reply of the Statistician for the Commission was to the effect that the control of accounts is not the control of financial policies; that the railroads were still free to make such disposition of their earnings as they chose: that "capital assets are consumed in operation . . . regardless alike of the corporation's policy of management or the Commission's rules of accounting"; that these rules deprived the railroads merely of the opportunity to misstate the facts.[1]

There is no reason to doubt the wisdom of the general policy expressed in these orders of the Commission. Charging betterments to operating expenses builds up a "secret reserve" which can be brought to light at any time and thus made as effective a means of manipulation as unearned dividends. The arbitrary diminution of the amount available as net income, by reason of swollen maintenance expenses, is unfair to the owners of income bonds or of non-cumulative preferred stock, or to other holders of equities in the business who may be more interested in maintaining present returns than in increasing future profits. Finally, it is not to the public interest that the existing level of rates should appear to be less profitable to the railroads than it really is.[2]

[1] *Twentieth Annual Report on the Statistics of Railways* (1907), pp. 21–24.

[2] "There is manifested [by the carriers] a growing appreciation of the

Some difficulties developed, however, in the handling of the depreciation accounts. The Commission did not prescribe the annual rates of depreciation which the carriers should use for different classes of equipment. It reported, "This latitude has resulted in widely different practices, the rates used by different carriers varying from nothing up to 7 per cent or more per annum upon equipment operated under substantially similar conditions. It seems evident that this variance is in some cases due to differences in policy rather than differences in actual depreciation or in the estimates thereof." [1] The result was, of course, that there were still many instances where the net income was either overstated or understated, according as the rate of depreciation used was too low or too high. The chief obstacle to the efficient control of depreciation charges has been the paucity of reliable statistics of the average life of different sorts of equipment under different operating conditions.

A more serious difficulty has arisen in connection with the disposition of the "reserve" which is created by the depreciation charges. Here we have to do with a matter of fundamental importance. The depreciation charges were inaugurated by requiring the carriers to include in their operating expenses a monthly charge of one twelfth of the estimated annual depreciation on each item of equipment. The determination of the various rates of depreciation to be assumed was, as we have seen, left to the carriers. An amount equal to the total charges for

importance of exact and theoretically correct accounting. At the same time there is a corresponding decline in the opinion, once too prevalent, that accounting may properly be governed by the policy or by the financial situation or needs of the accounting carriers." — *Twenty-Seventh Annual Report of the I.C.C.* (1913), p. 39.

[1] *Ibid.*

depreciation each month had to be credited to a "replacement" account. These replacement accounts were in turn charged with the cost of all new equipment acquired in place of equipment discarded or destroyed. Now it might happen, of course, that particular items of equipment would be worn out or destroyed before the accumulated depreciation charges on them would have amounted to enough to provide for their replacement. Even if the depreciation rates used had been found to accord with average experience there would have been a deficient provision for the replacement of all items of equipment which happened to be destroyed or abandoned before reaching the "average expectation of life."[1] Moreover, many items of equipment were already old on July 1, 1907 — the date at which depreciation charges began. So a "renewals" account for each class of equipment was also instituted. To the renewals account was charged as an operating expense the value of any item of equipment destroyed or otherwise retired from use minus salvage and the amount already accumulated in depreciation charges.

[1] Account should be taken of this official statement: "It is not intended that these accounts should be restricted to individual cars or locomotives. . . . On the other hand, the several amounts standing to the credit of these replacement accounts should be available to carriers for the purpose of replacement of equipment to the extent of such credits." — *I.C.C. Classification of Operating Expenses*, third revised issue (1907), p. 92. Later in response to a specific inquiry, it was definitely stated that the reserve created by depreciation charges on one class of equipment might be used for the purchase or replacement of other classes of equipment. (*I.C.C. Accounting Bulletin* no. 1, 1908, Case 108.) But these rulings are clearly inconsistent with the principle of the renewals account, by which provision was made for an adequate replacement reserve for each individual item of equipment. The real effect of the rulings was to make it possible for railroads to use the surplus in their replacement accounts in what were virtually additions and betterments, thus keeping their "property account" intact. The ruling in Case 108, referred to above, was later cancelled (*I.C.C. Accounting Bulletin* no. 4, 1909).

Thus, if a freight car which had cost $500 had been destroyed with a salvage of $75, when only $125 had been charged to depreciation for this car, a charge of $300 would at once be made to "renewals." The renewals account was a device for recording for each particular item of equipment the amount of finally realized depreciation not previously covered in the depreciation account. Operating expenses were in this way annually charged with (1) an amount representing the depreciation attributed to the wear and tear of the year's use, and (2) a further amount representing the total amount of depreciation on equipment retired within the year not previously written down.[1]

It seems to have been assumed that the amount of credits in the replacement reserve from accrued depreciation together with the amount of charges to renewals would stop the practice of charging to maintenance the expense of real improvements in equipment. At least in much of the current discussion of the problem it was assumed that depreciation charges would in some way be a gauge of proper allowances for maintenance. And it is hard to find any other significance in the official statement that "all replacements in excess of such credits must be considered as Betterments or Additions, and charged either to Income or to Capital." [2]

Whatever may have been the expectation, it was quite impossible that the magnitude of the replacement reserves should in any way have really determined the amount of expenditures properly chargeable to main-

[1] Subsequently the carriers were permitted to charge realized depreciation allocated to the period of use before July 1, 1907, to profit and loss instead of operating expenses. Cf. *I.C.C. Accounting Bulletin*, no. 8 (1912), Case 574.

[2] *I.C.C. Classification of Operating Expenses*, third revised issue (1907), p. 12.

tenance rather than to additions and betterments. For at any given time these reserves, together with the charges to renewals, would afford ample provision for all equipment retired from use and just about to be replaced, and would also contain *an additional amount equal to the accumulated depreciation charges on all equipment remaining in service.* This additional amount or surplus could not be written off for replacement purposes except as the particular items of equipment on which depreciation had accumulated went out of service,[1] and in the meanwhile it would in the normal course of things have been further swollen by the continuing accruals of depreciation charges. Since this continually growing surplus could not be used for legitimate replacement purposes, it furnished an opportunity to the carriers to charge expenditures for additional equipment to replacement and, hence, to operating expenses. The term "replacement account" was a misnomer.

When a more rigorous control of these matters was determined upon, those in charge of the accounting policies of the Commission had two possible methods of procedure open to them. One was to readjust their rules so that depreciation charges would in fact furnish an adequate replacement fund and nothing more. The alternative was to admit that the existing replacement accounts were not replacement accounts at all and to call them something else. This was done.

The new rule was phrased, "The accounts heretofore referred to as 'Replacement' accounts should hereafter be kept under the name of 'Reserve for Accrued Depreciation.'" [2] The total amount of the balances in these

[1] Despite official rulings to the contrary. Cf. note on p. 129, *supra.*

[2] *I.C.C. Classification of Expenditures for Additions and Betterments,* first revised issue (1910), p. 19.

accounts now had to be shown on the balance sheet as a deduction from the aggregate investment in road and equipment. This was a frank recognition of the permanent character of this reserve and of the fact that in the normal run of events it could not be diminished, but, on the contrary, was bound to increase. The account, moreover, had a new meaning. It was no longer supposed to have any definite bearing upon current replacements; it was held to represent the amount by which the values in the (equipment) property account had been diminished by the lapse of time. The regulation of expenditures for additions and betterments was accomplished in a simple and adequate fashion. It was required, as before, that when an item of equipment went out of service only the difference between its cost or ledger value and the subsequent accumulation of depreciation charges on it should be charged to operating expenses,[1] the remainder of the cost of replacement having already been charged to operating expenses through the depreciation account. But this was controlled by means of a new "equipment" account, the balance in which represented the difference between the aggregate expenditure for new equipment and the cost or ledger value of equipment retired. This balance was thus a measure of additions and betterments to equipment, provided that one measured replacements by the aggregate amount of realized depreciation (retirements) during the year. The equipment account, it will be noted, was quite dissociated from the accrued depreciation account, and in no way dependent upon depreciation charges of any sort.

[1] Specifically, to the "retirements" account, as the old "renewals" account was more appropriately renamed. The reserve for accrued depreciation was, of course, debited with an amount equal to the sum previously credited to it on account of the particular equipment retired.

The reserve for accrued depreciation was, then, identical with the old replacements account except that its significance was more accurately interpreted and that it was more adequately protected against being depleted to cover virtual additions and betterments. But to some carriers the change (in interpretation) seemed to carry with it large consequences. One inquiry reads: "This is a radical departure from the methods adopted by the Commission, July 1, 1907 . . . and the question now arises as to what carriers are to do with the vast sums they have charged in the past for depreciation on equipment." [1] Another is: "A carrier has left standing in its Replacement account a large balance, which will be constantly augmented, and it does not appear to be available for any purpose. How is this account to be closed?" [2] The accounting authorities of the Commission replied: "The effect of the methods prescribed for handling the accounts is that the account Property Investment will, in theory, include the actual cost of all equipment owned, while the depreciation account will represent the expired value or depreciation on the equipment." [3] This, of course, contemplated no closing of the account.

Let us turn from the Commission's regulations to the alternative procedure which it might have adopted when it was found that the replacement reserves were growing too rapidly to serve as a check on maintenance expenditures. If instead of changing the interpretation of the significance of the replacements account, it had desired so to modify the system of crediting to this account that it would have furnished a real gauge of proper

[1] *I.C.C. Accounting Bulletin*, no. 8, Case 567.
[2] *Ibid.*, Case 575.
[3] *Ibid.*, Case 567.

maintenance charges, this could conceivably have been accomplished.

Take the artificially simple case of a railroad owning n locomotives, each of which has cost $10,000 more than its scrap value and has a total period of serviceable life of ten years. Assume that the locomotives have been acquired in fairly even numbers in successive years to replace discarded equipment. Then of the n locomotives in service at the beginning of the fiscal year about one tenth will be discarded within the year. The normal depreciation charge per locomotive per year is $1000. So that if depreciation charges are inaugurated at the beginning of the year, the accruals for the year ($n \times$ $1000) will just suffice to replace the $n/10$ locomotives which will be discarded during the year. And in each succeeding year depreciation charges and replacement costs will continue to equal each other.

Up to the time when depreciation charges were begun the n locomotives then in service had completed in the aggregate $5n$ locomotive years. If the depreciation charges were to conform to the Interstate Commerce Commission's accounting rules it would have been necessary not only to make the provisions just described, but also to reach back into the past and set down depreciation charges on each locomotive retired for each of its years of life prior to the inauguration of the accounting system. And it would be necessary to proceed in this way until all of the original n locomotives had been replaced. Since the regular annual depreciation charges are adequate to provide for all replacements, the net result of the additional (retirements) charges would be the accumulation of a permanent and unusable reserve for accrued depreciation amounting to $5n \times$ $1000, or just half the difference between the original cost of the n locomotives and their scrap value.

In this simplified case current annual depreciation charges, without further additions on account of retirements, are a proper measure of current annual maintenance expenditures. Given sufficient data, there is no difficulty in thus equating depreciation and maintenance in the case of an actual railroad, with its equipment unevenly distributed in point of age. But the Interstate Commerce Commission's revised method of controlling replacement costs was simpler and better. The outcome of the equipment account was that the amount of annual expenditures for new equipment charged to operating expenses was determined by the (cost) value of the equipment retired within the year. Realized depreciation, rather than so-called partial depreciation (measured by the proportion of the aggregate probable life of the equipment on hand which has expired within the year) was the controlling factor.

In the case of a large railroad with varied equipment this completed depreciation (retirement) would be sufficiently uniform in amount, year after year, to obviate undue fluctuations in annual replacements. This method has the distinct advantages (1) that realized depreciation is a definitely measurable fact and (2) that replacements happen in fact to be made on account of realized depreciation and for no other reason.

By adequately providing for the control of expenditures for replacements this practice weakened the case for requiring depreciation accounts.[1] When these accounts were instituted the railroads had claimed that their properties were so large and varied, and their maintenance expenditures consequently so regular, that

[1] With the depreciation accounts eliminated, the entire cost of discarded equipment, minus salvage, would be charged to operating expenses under the rubric of "retirements."

maintenance itself was a sufficient protection against all depreciation. The rejoinder was, as we have seen, that in practice maintenance charges were frequently either too small to provide for replacements or were made unduly large through the inclusion of provision for additions and betterments. But with maintenance controlled, depreciation charges become, in effect, merely a way of measuring a presumed consumption of capital over and above the amount annually replaced.

Moreover, it was no longer possible to claim that to require depreciation charges was in no manner to interfere with the financial policy of the railroads. The compulsory annual additions to the reserve for accrued depreciation decreased by their full amount the apparent profits of the year and thus the amount available for dividends. Although no investment of a separate depreciation fund was required, yet the writing down of the capital assets by the amount of the accrued "depreciation" meant in the long run either that other assets had to be larger in amount than they otherwise would have been or that liabilities had to be smaller. Usually the growth of the reserve for accrued depreciation meant in practice that additional permanent investments were being made out of earnings. The reserve represented an additional, permanent, and compulsory investment in the business to take the place of the amount of the investment written off for depreciation. To require such a reserve to be created is to control, in that degree, the financial policies of the companies affected. This is not to condemn the requirement in question, but merely to show its real significance.

But, it may be objected, surely property that is half worn out is not worth as much as when it was new.

Should not its value be written down, if the balance sheets are to tell the truth? This question has significance only if it be assumed that the purpose of the depreciation accounts of railroads and other public service companies is to record the decline in the market value of particular assets. But even for private business undertakings this is not, in the opinion of the majority of competent writers, the purpose of depreciation accounts. Certainly it is not their purpose in railroad and public utility accounting.

The balance sheets of public service corporations are not designed, as those of a merchant might be, with reference to a possible insolvency. The outlook to be assumed is that of continuous and permanent operation. Many of the parts of a public service property could not be detached and sold except as scrap. And as for the market value of the undertaking as a whole, it is scarcely necessary to say that this is a matter of earning power, which depends upon rates and hence eventually upon the manner in which public control is exercised. It is hardly to be expected that any one will defend the propriety of a reserve against the possible depreciation of market values resulting from compulsory rate reductions.

Nor are the depreciation charges on public service properties supposed in any way to represent a diminution of the productive efficiency of the plant and equipment. With replacements and repairs properly attended to, there is no general decline in productive efficiency. Even for any individual item among the wasting assets the loss of productive efficiency usually comes as a sharp decline near the end of its period of life rather than a gradual deterioration spread evenly through its years of use.

The depreciation to be reckoned with in the case of such properties is merely a phase of cost-keeping — a device for allocating the consumption of capital among the successive years. The concrete facts to be recorded as best they may are not the "using up of values" (whatever that may mean) nor yet anything related to a deterioration in the service rendered, but merely the actual expiration of part of the aggregate probable period of use of the instruments of production on hand, and the nearer approach of the time when they will have to be replaced.

Profits and interest cannot be counted until the principal of an investment is replaced or provision is made for replacement. In an undertaking where one item forms the bulk of the income-yielding assets — a steamship, a coal mine, a patent right purchased at a price, for example — either enough must be set aside out of annual earnings to replace the original cost of the property when it is abandoned or retired, or the proprietors must be content to consider that their investment is being returned to them in installments. If the business is to be continued on the same scale the replacement must, of course, be provided for in advance or else the proprietors must make a second investment equal in amount [1] to the first.

All this is elementary. It is brought into the present discussion merely to indicate clearly the conditions under which a reserve for accrued depreciation is well-nigh indispensable. But where a property is varied and no single wasting asset or group of such assets is of dominating importance the case has been shown to be different. The periods of use of the different items of assets *overlap*, so that when depreciation is charged from the beginning

[1] Disregarding possible changes in the prices of capital goods.

on each item, a reserve begins to accrue on a given item before the reserves accumulated on account of other items have been diminished on account of replacements. The permanent reserve thus created is not needed. Nothing corresponding to it appears in the simpler case where depreciation is registered on some large asset representing the major part of the investment. It is very likely that some of its defenders fail to distinguish between depreciation charges to provide for replacement and depreciation to provide for liquidation — i.e., to maintain market value.

The fallacy in the view that the "reserve for accrued depreciation" is a necessary record of fact hinges on what is from the economic point of view the more or less accidental circumstance that the productive equipment of an undertaking happens to be in units of a sort that *are defined as units by the customary categories of purchase and sale.* In economic fact the property of a public service undertaking as a whole is a productive unit. Consider it as such — then replacements appear merely as repairs necessary to keep the whole property in a state of efficiency. Repairs in this large sense are of course to be counted as operating expenses, as is true of minor repairs. But if such repairs are fairly regular in amount year by year there appears to be no inexorable reason why a fund to provide for them should be accumulated in advance and more especially a fund that will provide more than the actual annual cost of the repairs.

But another and apparently weightier objection may be brought against this denial of the inevitableness of the depreciation reserve. If no such reserve is provided while the property changes from a condition of newness to a condition where it has on the average about half its

original expectation of life, does this not mean (if the business has been profitable) that about half the original investment has been returned to the pockets of the stockholders? And should not the amount of the investment as it appears in the balance sheet be written down accordingly? This point is apt to be stressed by defenders of the depreciation reserve.

These questions cannot be answered categorically. There is no necessary correlation between the mere aging or even the physical wear and tear of capital goods and the diminution of the investment. The concrete facts in the case are few. When capital goods are installed their cost is a definite amount of investment; when they are retired from use the investment is diminished by the amount of their cost, minus salvage. If such capital goods are replaced promptly when retired, is not the amount of the investment, in every real sense, kept intact?

The only assumption on which any other answer seems admissible is that in addition to repairs, replacements, and other costs, contributions to a reserve for accrued depreciation or to a replacement fund are *contemplated in advance as part of the expected costs of operation* and consequently taken into account in determining *the nature and amount of the original investment.* If the investment is made with this expectation, the failure to maintain such a reserve may properly be said to diminish it.

It can hardly be urged seriously that before the days of regulation by commissions a public service company ought to have decided to build up a useless "depreciation reserve." Against such a thesis we have the known facts (1) that such was not the general practice at the time, and (2) that the trend of court decisions was to the

effect that depreciation, even against replacement, need not be counted as part of operating costs.[1]

In the absence of such a reserve net profits for the time being [2] would of course have been higher than if a reserve had been accumulated. But to claim that this difference in the amount of profits represents the return to the proprietors of part of the principal of their investment is to beg the whole question. Such terms as "diminution of investment" and "repayment of principal" are clearly defined only in the case of a contractual loan. In the case of a permanent industrial investment for profit there is no such definition at hand. Lacking it, we can only turn to the expectations, plans, and estimates of the proprietors in order to determine what may properly be called net income.

To make the point clear we may take a marginal case. Suppose that an investment of $10,000,000 is made with the expectation of maintaining the property intact by all necessary replacements, but with no idea of accumulating a depreciation reserve. Suppose further that if contributions to such a reserve had, of necessity, entered into the calculations, the investment would not have been made, or, at any rate, would have been smaller. Can it be argued that the failure to maintain such a reserve is equivalent to repocketing part of the original investment?

In general, there is a reasonable presumption that the investments in undertakings which have not accumulated a depreciation reserve were not made with the expectation that it would be necessary to charge depre-

[1] See H. R. Hatfield, *Modern Accounting*, pp. 124, 125, and cases there cited.

[2] Until the properties had reached that static condition where replacements normally balance annual depreciation — and similarly for additions to the properties.

ciation accruals to operating expenses. It follows that it cannot be presumed that as a general rule the profits of such undertakings have contained an element which should be considered a repayment of part of the invested principal. Accordingly, objection might properly be made to a system of compulsory accounts which requires that property already on hand be written down for depreciation.

The control of accounting with respect to property acquired after the new system of accounts is introduced is quite another matter. These subsequent investments (including replacements) are made with full knowledge of the accounting rules in force and of the way in which "profits" are to be defined and measured. Whether a "reserve for accrued depreciation" should be required for such property is purely a matter of public policy. It is not necessary to the continued and successful operation of the property. But there are several points in its favor. Such inequalities as occur in the replacement needs of successive years can be met without shock. Annual profits tend to remain steadier when variations in replacements are absorbed by the variation of corresponding charges to the depreciation account. In the event of possible purchase at some future time by the Government the existence of adequate reserves against all accrued depreciation would greatly simplify the problem of valuation for sale, just as it would eliminate some of the difficulties that now attend the problem of valuation of rate control.

From the point of view of public policy, then, the depreciation rules of the Interstate Commerce Commission and of the various state commissions which have followed its lead are in general reasonable and apparently well-advised. They seem open to criticism only in so

far as they compel the accumulation of depreciation charges on property acquired *before* the rules were put into operation. Under the Interstate Commerce Commissions' regulations such charges are handled differently according as the depreciation is allocated to the period before or after July 1, 1907. Depreciation accruing subsequent to that date on property then in existence is, of course, charged to operating expenses. But this amount of depreciation is, as we have seen, roughly offset by replacements. It is only the depreciation charged on new additional equipment during its first years of use and on old equipment before the inauguration of the accounting system that accumulates as a permanent reserve.[1] It is only the practice of reaching back into the past for whatever depreciation is supposed to have accrued before July 1, 1907, that is open to objection.

But even in this respect the case is less serious than might be supposed. During the first two years of the new accounting régime, it is true, all of the realized depreciation (cost minus salvage) on equipment retired seems to have been charged directly to operating expenses. But under later rules that part of realized depreciation which was allocated to the period before July 1, 1907, was charged directly to profit and loss, and the companies had the option of charging such depreciation in the same manner before the equipment is retired. The effect of thus debiting these retroactive depreciation charges to profit and loss rather than to operating

[1] This is of course an arbitrary division of the depreciation charges. In fact all depreciation accruals over and above the amount of "realized depreciation" during a year accumulate as a reserve. The "realized depreciation" is in part on equipment installed before July 1, 1907. But there is, nevertheless, a rough equality between the depreciation realized in a year and the total amount of depreciation on equipment on hand credited to reserve during the year.

expenses was to write down the investment shown in the property account without providing for an equivalent increase in other assets. Instead the amount was deducted from the accumulated surplus (or added to the accumulated deficit) of the company. This does not affect annual earnings and it does not often affect dividends. This method of disposing of retroactive depreciation charges tended to bring the whole property account, so far as it represents wasting assets, into line with the new requirements.

But it will be many years before the property accounts of most American railroads will have much significance, apart from the significance attached to their annual fluctuations. For railroads, as for lesser public utilities, the sum of permanent investments on which earnings are to be permitted can rarely be ascertained from the printed balance sheet. This being true, there is no reason why either the Commission or the courts should consider these new accounting devices as throwing any light on the amount of the investment up to July 1, 1907. The retroactive part of the depreciation charges should have no bearing upon the valuation of railroad properties. In the long run there will be no substantial injustice to the railroads so long as the depreciation accounts remain merely a device for helping to determine what portion of gross earnings shall be available for distribution as net earnings.

This brings us back to our main problem. The significance of our digression into the matter of accounting regulations should now be apparent. To diminish the valuation of a railroad or other public service property on account of depreciation is to assume that depreciation should have been charged on each item of plant and

equipment from its first installation. If the accounts of the company had been subject to regulation from the beginning, and if the accounting regulations had provided for depreciation charges, no exception could be taken to the recognition of depreciation as an element to be taken into account in valuation. But in such a case valuation would be unnecessary. The accounts would tell the story. Lacking such accounts we turn, perforce, to the value (there is no better word) of the tangible properties as affording about the only available evidence of the amount of investment entitled to a return. If the company has not set apart some of its annual earnings on account of depreciation, shall we assume that it should have done so, and scale down the value of the property accordingly?

With the exception of the St. Louis Public Service Commission all of the commissions that have squarely faced this question have answered in the affirmative. So has the United States Supreme Court. But the St. Louis Commission, in the cases previously cited,[1] held that if replacements and renewals have been properly attended to, no deduction for depreciation should be made. To assume that a company should have provided for a depreciation reserve over and above its provision for replacements is, in effect, to assume that the absence of such a provision means that some of the investment has been returned to the stockholders along with their ordinary profits. To inquire into this last matter, said the St. Louis Commission, is virtually to regulate past profits — and that task it refused to attempt.

This conclusion of the St. Louis Commission is, of course, in line with the general results of our analysis

[1] *Supra*, p. 121, note.

of depreciation charges. The case for the position of the St. Louis Commission seems to be stronger, however, if we put aside the difficult questions suggested by the reference to the regulation of past profits, and again recall (1) that "income" and "repayment of principal" are not, in fact, easily separable, and (2) that the expectations of the investors as to the expense of maintaining the invested capital intact may properly be taken as a criterion, provided that their plans provide adequately for maintaining the productive efficiency of the plant and equipment and are in harmony with prevailing business practice and with the law.[1]

But the contrary view is the usual one. The opinions of the Wisconsin Railroad Commission may be taken as representative, since that body was a pioneer in this field and its precedents have been largely followed by other commissions. In one of its leading cases it has held as follows:

> Depreciation may be described as the amount that must be regularly set aside to cover wear and tear, etc., in order to keep the original investment intact. It is an operating expense and should be borne by the customers through the rates paid by them for the services rendered by the utility. But when depreciation is so borne by them, it should be set aside until needed for the renewal of worn out or useless parts of the plants. If under these conditions it is not so set aside and used, but diverted to the stockholders for their use or personal benefit, this diversion is tantamount to the payment of dividends out of the capital. It simply means that the money contributed by the consumers for the upkeep of the plant and the investment has been paid over to the stockholders

[1 In other words, the reasonable expectations of the planners of an enterprise with respect to the practice which they should follow in providing for replacements must be considered to have entered into the supply price of the capital attracted into the enterprise.]

> instead of being devoted to the purposes for which it was properly intended. It can mean nothing else. Since depreciation, in a sense, is intended to keep the investment intact, it necessarily follows that by turning it over to the stockholders, a part of their capital is in reality returned to them, and that this, in turn, is reducing their investment in the plant. Since their investment is thus reduced, it would also seem that there should be corresponding reductions in the amount upon which the rates paid by the consumers are based. There would certainly seem to be instances where no other course would be equitable all around, unless the capital that has thus been withdrawn by the stockholders is restored to the depreciation fund. Investors no more than any one else can both eat their cake and have it.[1]

This argument begs the question. It assumes that unless a useless depreciation reserve is accumulated, money contributed by consumers has been put into the pockets of stockholders as a virtual repayment of some of the capital originally invested. The Commission was confusing depreciation for replacements with physical depreciation as measured by the lapse of time. Its opinion was that "as depreciation is constantly going on, the charges by which it is covered should also be regular. That is, a sufficient amount of money should be set aside each year to cover the cost of replacing each part of the plant as it becomes useless or unfit for further use." Evidently the Wisconsin Commission regarded depreciation sufficient to cover replacement as entirely adequate. It failed to see that if a public service property has been properly maintained, all depreciation necessary to replacements is automatically absorbed by replacements and renewals themselves, and that the further accumulation of a depreciation reserve is a

[1] Hill *v.* Antigo Water Co., 3 W.R.C.R. 641.

burden which it is neither necessary nor just to assume should have been self-imposed by the companies.

Further inconsistencies are found in other opinions of the Wisconsin Commission. Thus we find in the case of State Journal Printing Co. *v.* Madison Gas and Electric Co.:[1] "In the long run the total renewals should amount to the total depreciation, but for any given period there may be wide differences between them. During the earlier years of the life of the plant the depreciation is likely to be the greater. During some at least of the subsequent years the renewals are apt to be the greater." In another case[2] the same Commission said, "the depreciation reserve should ordinarily be sufficient to provide for wear and tear, obsolescence, and inadequacy," but in their valuations the Commission insists that the depreciation reserve shall in effect be larger than this — larger in fact by practically the whole amount of the sum deducted from the valuation on account of "depreciation."

In the decisions of the Supreme Court we find precisely the same inconsistencies. In its first definite pronouncement on this point[3] the Court held: "The cost of reproduction new is one way of ascertaining the present value of a plant like a water company, but that test would lead to obviously incorrect results, if the cost of reproduction is not diminished by the depreciation which has come from age and use." But it is important to know the reasoning on which this conclusion rested. This is given in another part of the opinion:

> A water plant, with all its additions, begins to depreciate in value from the moment of its use. Before coming

[1] 4 W.R.C.R. 560.
[2] King *v.* Wisconsin Telephone Co., 10 W.R.C.R. 521.
[3] Knoxville *v.* Knoxville Water Co., 212 U.S. 7 (1909).

to the question of profit at all the company is entitled to earn a sufficient sum annually to provide not only for current repairs but for making good the depreciation and replacing the parts of the property when they come to the end of their life. The company is not bound to see its property gradually waste, without making provision out of earnings for its replacement. It is entitled to see that from earnings the value of the property invested is kept unimpaired, so that at the end of any given term of years the original investment remains as it was at the beginning. It is not only the right of the company to make such a provision, but it is its duty to its bond and stockholders, and, in the case of a public service corporation at least, its plain duty to the public. If a different course were pursued the only method of providing for replacement of property which has ceased to be useful would be the investment of new capital and the issue of new bonds or stocks. This course would lead to a constantly increasing variance between present value and bond and stock capitalization — a tendency which would inevitably lead to disaster either to the stockholders or to the public, or both. If, however, a company fails to perform this plain duty and to exact sufficient returns to keep the investment unimpaired, whether this is the result of unwarranted dividends upon over-issues of securities or of omissions to exact proper prices for the output, the fault is its own. When, therefore, a public regulation of its prices comes under question the true value then employed for the purpose of earning a return cannot be enhanced by a consideration of the errors in management which have been committed in the past.

So far as this means that a company should not be allowed to add the cost of replacements to its capital account, it would command general assent. But so far as it means that a "reserve for accrued depreciation" should have been maintained or (what comes to the same thing) that public service property should be

"depreciated" in a valuation, it defeats its own purposes. The gist of the argument is that depreciation should have been counted from the beginning in order to provide for the "replacements of property which has ceased to be useful." The Supreme Court has fallen into the same error of *fact* which entrapped the Wisconsin Commission.

For simplicity's sake I have considered only that common and universal sort of depreciation which is measured by the approach of the time when, in normal course, replacements will have to be made. The public service properties which I have had in mind are sufficiently large and varied to give validity to the assumption that depreciation accruals and replacements come in fairly regular amounts in successive years.

There are many cases which these general assumptions do not fit and for which our conclusions have to be modified. In a small railroad the cost of a certain bridge may be disproportionately large, and it may be impossible to replace it from earnings except by annual accumulations. In valuing such a railroad an allowance should be made for the depreciation of the bridge. So with small public utilities for which the power plant or some other wasting asset may be of dominating importance. But the principle of such exceptions is simple.

Nothing has been said about "depreciation for obsolescence." Obsolescence is unlike depreciation by reason of age because it cannot be foretold in advance — even as an average for a group of items. Obsolescence, then, is a matter of fact, not of formula. Any reserve established to cover losses on account of obsolescence is really a special contingency reserve — closely analo-

gous to an insurance fund. Whether allowance should be made for obsolescence in valuation procedure is a different question from the one we have been considering. [1]

Summarized, our conclusions then are:

1. If depreciation charges have not been required by public authority, it cannot be assumed that the proprietors of a large public service undertaking should have accumulated a reserve for accrued depreciation.

2. The absence of such a reserve does not necessarily mean that part of the principal of the investment has been returned to the proprietors.

3. In valuation for purposes of rate control no deduction should be made on account of the depreciation of large and varied properties, except for depreciation allocated to a period in which depreciation accruals were regularly charged to operating expenses.

[1] The relative merits of different methods of apportioning depreciation among the years of life of an item of equipment have also been passed over. For convenience it has been assumed that depreciation charges, if imposed at all, are to be apportioned by the "straight-line method" (equal annual increments). The argument is independent of this assumption.

IX

THE SHERMAN ACT AND THE NEW ANTI-TRUST LEGISLATION [1]

I

THE recent additions to the federal legislation having to do with monopolies and trusts have followed upon three years of continued and vigorous discussion. In part this general revival of interest in the control of the trusts was bound up with the growth of a general radical movement in politics. But it is not mere coincidence that this period of active public discussion followed immediately upon the decisions of the Supreme Court in the Standard Oil and American Tobacco cases. Before these decisions the Sherman Act was a weapon which had never been fairly tested. With its strength and its limitations partly uncovered it became a fair target for criticism.

Each of the undertakings directly affected by these decisions was precisely the sort of undertaking against which the statute was primarily directed. Whether labor organizations, railroad combinations, and price agreements affecting a limited area were illegal might afford ground for debate, but the intent of the statute in regard to great industrial combinations was clear. Not until the law had been on the statute books for over twenty years, however, was its efficacy as a weapon against unified industrial consolidations — "trusts" of the modern sort — thoroughly tested in the Supreme Court.[2]

[1] Reprinted, with some omissions, from the *Journal of Political Economy*, vol. XXIII, nos. 3, 4, and 5 (March, April, May, 1915).]

[2] The E. C. Knight Co. (Sugar Trust) case does not constitute a real

Furthermore, the findings of the Court were such as to stimulate discussion. Especially is this true of the Court's intimation that not all agreements in restraint of trade were prohibited. Many persons found this state of affairs in itself unsatisfactory. Others were disquieted by the opportunity for judicial legislation afforded by the necessity of drawing a line between legal and illegal combinations. There was also an excusable curiosity on the part of business men as to the legal status of existing business arrangements, including not only overt combinations, but also a mass of secret trade agreements which would be brought to light, in part at least, if their legality were assured. There was a similar interest in the possibility of introducing trade agreements into new and supposedly forbidden fields.

There was also an increased amount of apprehension, especially on the part of the larger business combinations. Even if the law itself had been somewhat softened, its administration had been unprecedentedly vigorous. There have been some attempts to show that the vigor with which the law has been enforced has varied with the policies of different presidential administrations. Since the law was enacted there has been a rapid but fairly steady increase in the number of prosecutions and a yet more notable increase in the number of successful prosecutions. But it would be going too far to attribute all of the difference to the apathy of the earlier administrations. The earlier decisions, especially in the cases of *In re* Greene [1] and United States *v.* E. C.

exception to this statement, because of the narrow issue on which the decision of that case turned. None of the other so-called "trusts" involved in cases which reached the Supreme Court were more than combinations of independent undertakings held together by price agreements, pooling arrangements, or the employment of a common sales agent.

[1] 52 Fed. 104.

Knight Co.,[1] were discouraging, and the development of adequate administrative machinery has been a slow process. At any rate, it is evident that in recent years the work of the public prosecutor has been carried on under more advantageous conditions than it was twenty or ten years earlier. Successive decisions have indicated what are likely to be the more vulnerable aspects of the combinations attacked; mistakes and deficiencies in the preparation of some of the earlier cases have been guarded against in the later. Moreover, whether or not a given combination falls under the condemnation of the statute is a matter of fact as well as of law, and the Department of Justice is much better equipped than it was for uncovering the facts. Not until 1903 [2] was the immunity of witnesses from prosecution on account of matters testified to, which had been granted in 1893 so far as proceedings under the Interstate Commerce Act were concerned, extended to witnesses in proceedings under the Anti-Trust Act. The creation of the Bureau of Corporations in 1903 and the organization of the Bureau of Investigation in the Department of Justice in 1908 [3] made a permanent staff of expert investigators available. These bureaus have been signally useful in the collection of evidence for the prosecution, and what is quite as important, their investigations have indicated in advance whether a proposed prosecution under the statute might be undertaken with any hope of success.[4] The Government has been better equipped in recent years for the effi-

[1] 156 U.S. 1.

[2] 32 U.S. Stat. 854. Mention should also be made of the Acts of 1903 and 1910 to expedite hearings under the Sherman Law (32 U.S. Stat. 823; 36 U.S. Stat. 854).

[3] *Report of the Attorney-General*, 1909, p. 8.

[4] *Ibid.*, 1910, pp. 2, 3, 26.

cient enforcement of the Sherman Law than ever before.

But it would not be fair to attribute all of the recent interest in the amendment of the Sherman Act to dissatisfaction and unrest created by the interpretation and administration of the statute. Legislators and publicists had a more adequate appreciation of the problems involved in the growth of industrial combinations than they had twenty years before. That the Sherman Law had apparently had but little effect in checking the movement toward industrial combination was in itself a fact demanding explanation.

It is significant that in much of the more serious discussion, certain principles were recognized that for some time had been generally accepted among economists. In fact, there have been specific instances of the direct influence of economic writing and teaching.[1]

There seems to be a very general impression that the decisions in the Oil and Tobacco cases declared a new judicial policy; that "good trusts" were to be distinguished from "bad trusts"; that combinations might restrain trade if they did not restrain it "unduly"; that possibly even price agreements were permissible, if the prices agreed upon were "reasonable." Much has been written upon these points, but because the significance of the recent supplementary legislation hinges upon the real bearing of these decisions, I think it worth while to review the matter again.

[1] Special reference should be made to the influence of (1) Professor R. T. Ely's early and continued insistence upon the fact that mere size does not give sufficient advantage to be the basis of a lasting monopoly, and that where there is such monopoly there must be some definite source of monopoly power, and (2) Professor J. B. Clark's thesis that the only power for evil possessed by most of the trusts was the power of predatory competition.

If those decisions are to be taken at their face value, the general impression that they opened the way for "reasonable" price agreements and mildly repressive contracts and combinations in restraint of trade does not seem to have been well founded. The precedents from which the new holdings were supposed to mark a departure were established in the Trans-Missouri Freight Association and Joint Traffic Association cases. In those cases it was held in general terms that the Sherman Law, referring as it does to "every" contract, etc., in restraint of trade, prohibits contracts and combinations in restraint of trade whether reasonable or not. The minority opinion in the Trans-Missouri case, written by Mr. Justice White and concurred in by three other justices, held that the phrase "restraint of trade" should be given its common-law meaning, under which not all contracts limiting the freedom of trade are deemed invalid as involving "restraint of trade."

But the issue as between the majority and minority opinions was not squarely joined. It has been generally overlooked that it was intimated in the majority holding in the Trans-Missouri case and plainly stated in the Joint Traffic case that certain kinds of contracts which really involve some restraint of trade but are not invalid under common law would not be held illegal under the statute.[1] From this it would appear that so far as the general statement of the principles of the law is concerned the division of the Court was more apparent than real. But there was, nevertheless, some fundamental ground of disagreement, for these supposedly conflicting interpretations of the law were not offered as mere *dicta*, but in support of opposing opinions as to the legality of the particular combinations under consideration.

[1] 166 U.S. 329; 171 U.S. 567.

The specific question was whether a combination of railroads for the purpose of fixing and maintaining rates fell under the condemnation of the statute, if the rates fixed by the agreement were not shown to be unreasonable. In the dissenting opinion it was claimed, first, that such agreements among railroads were necessary and even desirable, hence not unreasonable and consequently not illegal, and, second, that the Sherman Act was not intended to apply to railroads. The majority of the Court held that neither of these contentions could be sustained in the face of the fact that the Sherman Act condemns *every* combination in restraint of trade. Although both the majority and the minority of the Court agreed that some agreements which restrain trade are not condemned by the statute, they did not agree as to the basis of such exceptions. The dissenting opinion held that "reasonableness" in the sense, apparently, of an absence of harmful economic or social effects is the ground on which exceptions should be made. The majority of the Court held (more clearly in the Joint Traffic case) that every agreement is illegal if its direct purpose is restraint of trade, while it may be legal if the restraint of trade is merely incidental to a contract made primarily for another and legitimate purpose, providing that the restraint of trade is no greater than is necessary to the protection of the legitimate purpose of the contract.[1] Both the majority and minority opinions claimed to be in harmony with common-law precedents. But this last is a matter too large for present discussion, and does not fall within the competence of the present writer.[2]

[1] 171 U.S. 567, 568. Cf. Hopkins *v.* United States, 171 U.S. 592–94.

[2] The majority opinion seems to be in harmony with the frequently cited summary of common-law precedents by Judge Taft in United

What the Trans-Missouri and Joint Traffic decisions virtually resolve themselves into is this: Certain classes or forms of agreements, such as a partnership or an ordinary corporation formed by the union of former competitors, or the sale of the goodwill of a business on terms by which the seller agrees not to compete with the buyer, may "restrain trade," but because the main purpose of such combinations or contracts is legitimate, and the restraint of trade is only incidental, they will not be considered as restraints of trade under the statute. But where the *direct purpose* of a combination or contract is restraint of trade it will be deemed illegal even if the restraint is reasonable in the sense that it is not prejudicial to the public interest. The things thus distinguished have come to be known as "direct restraint" and "indirect restraint."

The Court did not differentiate the meaning of "restraint of trade" in the statute from its common-law meaning by finding that any of the general classes or kinds of contracts or combinations which were valid under common law were illegal under the statute. But it did refuse to give to the statute the elasticity which the dissenting opinion held was to be found in the common law. Congress, in the view of the court, had established unrestricted competition as public policy, and it was for Congress, not the Court, to relax the rule. In short, all combinations and contracts entered into with the direct purpose of restricting competition in interstate commerce were held to be illegal, although it was

States *v.* Addyston Pipe & Steel Co., 85 Fed. 281, 282. In general harmony with the dissenting opinion is Mr. Bruce Wyman's thesis that the public interest rather than the hard-and-fast rule of the maintenance of competition is the ultimate criterion by which the common law tests the validity of contracts in restraint of trade. Cf. his *Control of the Market*, especially chap. vi.

still held, as in earlier cases, that agreements which "indirectly and remotely" affect interstate commerce were not covered by the act.

The doctrine thus established was followed in the Northern Securities case [1] and in a number of cases decided in the inferior courts.[2] In none of these subsequent cases, however, with the possible exception of the Northern Securities case, is there any clear indication that the findings would have been for the defendants had it not been for the Trans-Missouri precedent. The rule established in that case may merely have afforded the courts a convenient means of narrowing the issues by ruling out at once the defense of reasonableness.

On the other hand, however, before the Trans-Missouri decision the lower courts had refused to apply the Sherman Act to a number of combinations which clearly restrained trade in ways more direct than were indicated as permissible in the Trans-Missouri case. Certain price agreements, for example, were declared legal.[3] After the Trans-Missouri decision, agreements somewhat similar to these were declared unlawful.[4] It is reasonable to infer that the Trans-Missouri decision really widened the actual field of application of the statute.

Turning now to the Standard Oil and American

[1] 193 U.S. 331.

[2] Representative cases are: United States *v.* Coal Dealers' Association, 85 Fed. 252; United States *v.* Swift & Co., 122 Fed. 529; Whitwell *v.* Continental Tobacco Co., 125 Fed. 454; Phillips *v.* Iola Portland Cement Co., 125 Fed. 593.

[3] United States *v.* Nelson, 52 Fed. 646; Dueber Watch-Case Manufacturing Co. *v.* E. Howard Watch & Clock Co., 66 Fed. 637.

[4] Ellis *v.* Inman, Poulsen & Co., 131 Fed. 182; Loder *v.* Jayne, 142 Fed. 1010; Dr. Miles Medical Co. *v.* John D. Park & Sons Co., 164 Fed. 803; 220 U.S. 373.

Tobacco cases, is there any clear evidence that the Court's expressed intent to interpret the statute "in the light of reason" meant that its field would again be restricted? No definite answer can be given except in the light of subsequent decisions in border-line cases. The *dictum* in the Standard Oil case does not afford a satisfactory basis for inference.

Taking the decision at its face value, however, it does not seem to come to quite the same thing as the minority opinion in the Trans-Missouri case, although both were written by Chief Justice White. In that dissenting opinion it was held without equivocation that only "unreasonable" contracts in restraint of trade were condemned by the statute, and it was urged that to construe the statute as excluding certain classes of contracts, such as those which are collateral to a sale of property, while holding at the same time that the statute covers every contract in restraint of trade, was illogical. Had this minority opinion prevailed, it would have given to the statute all the elasticity of the common law. The courts would have been free to pronounce upon the legality of particular kinds of combinations according to their view of public policy. The determination of public policy would have been influenced, of course, by precedents, but many modern forms of combinations are too new and precedents too uncertain to make possible any hard-and-fast rule of discrimination between permissible and illegal combinations. For better or worse, the road to judicial legislation would have been opened.

The Standard Oil decision did not state explicitly that only "unreasonable" contracts in restraint of trade were condemned by the law. Its emphasis was on the thesis that the statute must be interpreted "in the light

of reason," and that the "rule of reason" must be applied in determining whether a particular combination is within the scope of the law's condemnation. It is true that "undue" restraint of trade was specified as the thing prohibited, but the stress was not put on this point. There is nothing in the decision really inconsistent with the minority opinion in the Trans-Missouri case. It went farther than that holding, however, in that it *identified restraint of trade with monopoly*.

It was pointed out that with the gradual narrowing of the older common-law definition of restraint of trade as any voluntary restraint put by an individual on his right to carry on his trade, and the accompanying broadening of the older meaning of monopoly so as to include engrossing, the notions of restraint of trade and of monopoly came to be blended. Interpreting the statute, then, as using common-law phrases with a common-law meaning, it was held that its first section, declaring contracts, combinations, and conspiracies in restraint of trade illegal, and its second section, making it illegal to monopolize or to attempt to monopolize any part of interstate trade or commerce, are mutually explanatory. "Undoubtedly," said the Court, "the words 'to monopolize' and 'to attempt to monopolize' as used in the section reach every act bringing about the prohibited results."

If this joining of the first two sections of the act could be accepted as final, the definition of the word "monopolize" would be the only serious difficulty in the interpretation of the statute. The courts would be left free to determine whether a given restraint of trade is against public policy, but the standard by which public policy should be gauged would be definitely prescribed. In certain respects this may be taken to be the actual state

of the law. The particular form which a combination adopts counts for nothing one way or the other. Its purpose and the methods used in attaining its purpose are the crucial facts. If "judicial legislation" means the determining of public policy in particular instances by the courts, the Standard Oil decision did not give it much scope.

Yet this is not all there is to the matter. There are important classes of cases to which the Sherman Act has been held and is still held to apply in which the presence or absence of monopoly or of monopolistic intent is *not* the determining factor. Under the ruling interpretations of the statute, there may be "restraint of trade" where there is no "restraint of competition." Most of these cases have to do with interference with commerce on the part of labor unions.

In the Sherman Act the common-law phrase "restraint of trade" became "restraint of trade or commerce among the several states, or with foreign nations." This and the fact that the act rests upon the federal power over interstate commerce have given a decided twist to its interpretation. The contracts in restraint of trade found in the early common-law cases were agreements by which a person bound himself to refrain from pursuing a specified calling. The evils involved were supposed to reside partly in the partial loss of the power of supporting one's self and of contributing to the welfare of the community. Later, agreements tending to limit or suppress competition as an efficient price-fixing force were also classed as contracts or combinations in restraint of trade. But, so far as I can find, the effect of such agreements upon the *volume of commerce* was never made the factor determining their illegality.

In one of the early cases under the Sherman Act [1] it was held by the Circuit Court for the District of Massachusetts that the words "commerce" and "trade," as used in the statute, are synonymous:

> Careless or inapt construction of the statute . . . will, if followed out logically, extend into very large fields, because . . . the inevitable result will be that the federal courts will be compelled to apply this statute to all attempts to restrain commerce among the states, or commerce with foreign nations, by strikes or boycotts, and by every method of interference by way of violence or intimidation. It is not to be presumed that Congress intended thus to extend the jurisdiction of the courts of the United States without very clear language.[2]

But within a month another circuit court had granted an injunction against a combination of striking workingmen on the ground that in refusing to work they were violating the statute by interfering with the movement of goods in interstate commerce.[3] The opinions in the cases arising out of the railway strike of 1894 contain explicit recognitions of their departure from common-law precedents. Thus, "The term 'restraint of commerce' was used in its ordinary, business understanding and acceptation." [4] Another statement is:

> I am unable to regard the word "commerce," in this statute, as synonymous with "trade" as used in the common-law phrase "restraint of trade." In its general sense, trade comprehends every species of exchange or dealing, but its chief use is "to denote the barter or purchase and sale of goods, wares, and merchandise, either by wholesale or retail," and it is so used in the phrase mentioned. But "commerce" is a broader term . . .

[1] United States *v.* Patterson, 55 Fed. 605.
[2] 55 Fed. 641.
[3] United States *v.* Workingmen's Amalgamated Council, 54 Fed. 994.
[4] United States *v.* Elliott, 64 Fed. 30.

> and, as used in this statute. . . . should not be given a meaning more restricted than it has in the Constitution.[1]

Even more specific is the following:

> The primary object of the statute was, undoubtedly, to prevent the destruction of legitimate and healthy competition in interstate commerce by individuals, corporations, and trusts, grasping, engrossing, and monopolizing the markets for commodities. But its provisions are broad enough to reach a combination or conspiracy that would interrupt the transportation of such commodities and persons from one state to another.'[2]

For present purposes it is immaterial whether the logic of these opinions is sound, or whether the Sherman Act was seized upon and distorted somewhat in order to meet an emergency.[3] But even if these interpretations of the scope of the act were not illogical, they were in no sense necessary. Whether the statute was intended to apply to labor combinations may be a matter of doubt, but there is no evidence in the congressional debates that it was anticipated that the statute would prohibit interference with commerce as a thing distinct from restraint of trade. The opinions make much of the use of the term "conspiracy" in the statute, but even if that word means in this connection anything more than "combination," [4] the decisions nevertheless

[1] United States *v.* Debs, 64 Fed. 749.

[2] United States *v.* Cassidy, 67 Fed. 705.

[3] There may be some significance in the fact that the Supreme Court, in the only one of these cases which reached it (*In re* Debs, 158 U.S. 564), did not rest its findings upon the Sherman Act, but on the general federal powers over interstate commerce and the transmission of the mails. But its subsequent opinions have been such as to imply an approval of the broad scope given to the Sherman Act by the lower courts in the railroad strike cases.

[4] Justice Holmes held in his dissenting opinion in the Northern Securities case (193 U.S. 403) that "the words hit two classes of cases, and only two — contracts in restraint of trade, and combinations or conspiracies in restraint of trade." Cf. Rice *v.* Standard Oil Co., 134 Fed. 464.

really hinge upon the meaning attached to the word "commerce."

Similar in some respects are the decisions of the Supreme Court in two cases that should be sharply differentiated from most of the cases under the statute that have come before that body. In Loewe *v.* Lawlor[1] a boycott instituted by the Danbury Hatters' Union was held to constitute a conspiracy in restraint of interstate commerce. It was immaterial, said the Court, that there was no suppression of competition, and it was even immaterial that there was no direct physical obstruction to the movement of commerce. It was sufficient that the boycott tended to check the free flow of goods from one State to another. In the case of Eastern State Lumber Association *v.* United States[2] an arrangement by which the members of an association of retail lumber dealers were informed of the names of wholesale dealers who sold directly to consumers was held to be a similar restraint upon the movement of interstate commerce, although no suppression of competition was alleged.

Doubtless in many jurisdictions the actions condemned in these two cases would have been held to be offenses under the common law, but not because they "restrained trade" and certainly not because they inter-

[1] 208 U.S. 274.

[2] 234 U.S. 600. The older case of Montague & Co. *v.* Lowry, 193 U.S. 38, turned partly upon similar considerations, although in that case the element of monopoly was not entirely absent. And in United States *v.* Patten, 226 U.S. 541, the Court said: "Sec. 1 of the act, upon which the counts are founded, is not confined to voluntary restraints, as where persons engaged in interstate trade or commerce agree to suppress competition among themselves, but includes as well involuntary restraint, as where persons not so engaged conspire to compel action by others, or to create artificial conditions, which necessarily impede or burden the due course of such trade or commerce or restrict the common liberty to engage therein."

fered in any way with the general movement of commerce.

The fact seems to be that the Sherman Act is held to apply to two very different classes of things. Interference with the movement of commerce is one thing and suppression of competition is quite another thing. In point here is the extraordinary decision of the Circuit Court of Appeals in Whitwell *v.* Continental Tobacco Co.,[1] which held that under a correct interpretation of the second section of the act "no attempt to monopolize a part of commerce among the states is made illegal or punishable unless the necessary effect of that attempt is to directly and substantially restrict commerce among the states." Probably this decision (which refused to see in an attempt to monopolize by the use of a factor's agreement of the most flagrant sort any evidence of the violation of the act) would not now be accepted as good law. But it draws a sound distinction between suppressing competition and hindering the movement of commerce. Few economists would care to maintain that competition always has the advantage over monopoly in respect of the volume of commerce it creates. It seems then that the identification of "restraint of trade" and "monopolizing" or "attempting to monopolize" is not complete. The Sherman Act, as at present interpreted, condemns *two different sorts* of offenses.

The Sherman Act is a general statute, declaratory of public policy. As such it must be judged by (1) the soundness of the public policy which it declares, (2) the accuracy and completeness with which it declares that public policy, and (3) the adequacy of the mechanism which it provides for making that policy effective.

[1] 125 Fed. 454.

(1) The public policy which the act was intended to embody is that competition should be maintained, artificial monopoly destroyed, and its growth prevented. In the congressional debates attending its enactment the great industrial combinations of the day were singled out by name as the things against which it was directed. It is clear from the debates that the hostility toward these industrial combinations was especially directed against (*a*) their supposed power over prices and (*b*) their aggressive suppression of competition. Whatever the economic advantages of monopoly *per se* may be, there will be little question of the soundness of the policy which would attempt to deprive it of its power for evil in these two particulars.[1]

Sometimes it is urged that to attempt to maintain competition by prohibiting attempts to monopolize is illogical, since monopoly is the goal of competition, and achieved monopoly is merely the result of thoroughly successful competition.[2] The aim of each competitor is

[1] The more weighty discussions of the economic advantages of monopoly have to do with the effect of monopoly upon the aggregate production of wealth. From this point of view the case for monopoly is exceedingly dubious and, at best, has a validity that is restricted and conditioned in many ways. And such considerations are relatively unimportant compared with matters like the effect of monopoly upon distribution, upon the scope for individual initiative, upon economic opportunity in general, and upon a host of social and political relations.

[2] Such was the opinion of a number of witnesses before the Senate Committee on Interstate Commerce, 62d Congress. See for example the Hearings pursuant to Sen. Res. 98, pp. 1431 ff. One of the few expressions of similar views by economists is to be found in Professor R. L. Liefmann's article entitled "Monopoly or Competition as the Basis of a Government Trust Policy," *Quarterly Journal of Economics*, vol. xxix, p. 308 (February, 1915). For a judicial opinion to the same effect, see Whitwell *v.* Continental Tobacco Co., 125 Fed. 462, where it was even held that "every sale and every transportation of an article which is the subject of interstate commerce is a successful attempt to monopolize that part of this commerce which concerns that sale or transportation." In general, however, this abuse of the word "monopolize" has no better standing in law than it has in economics.

to aggrandize his own business at the expense of his rivals, and just so far as he leaves his rivals in partial possession of the field he falls short of complete success. The simplicity of this reasoning makes it attractive, but it is too simple to fit the facts. We need not stop to discuss the pertinent question as to whether a permanent industrial monopoly can be achieved by the use of permissible competitive methods by a business not enjoying any special advantages not open to its competitors aside from the possession of a larger amount of capital, for the matter can be disposed of in a simpler fashion. There is a difference between competing and "attempting to monopolize." The ordinary competitor does not have monopoly in mind as an end. He strives to increase his profits by increasing his trade, and he endeavors to acquire as much as he can of the custom enjoyed by his competitors. But the thing directly sought is such increase of his profits as is possible under competitive conditions. The effect of his policies on his competitors is secondary and, generally speaking, indirect. The phrase "attempting to monopolize" on the other hand is meaningless if it does not refer to definite efforts to get rid of the limitations which competition sets upon one's ability to buy and sell at such prices and on such terms as one pleases. The elimination of competition through the absorption or crippling of competing establishments becomes the direct and primary object, and the methods used are adapted to this end. The contention that "to compete" and "to attempt to monopolize" are synonymous is unsound. They are antagonistic in principle.

(2) Is the Sherman Act an accurate expression of the public policy which it seeks to declare? If by accuracy is meant precision, it has little of it. It is a lawyer's

statute, speaking the language of the common law. It was drafted by the Judiciary Committee of the Senate as a substitute for the Sherman Bill, which, together with a mass of proposed amendments, had been referred to that committee. These amendments included various lists of particular offenses and various delimitations of the field of application. It was evident that it would be difficult for Congress to come to an agreement on any one set of particularizations. The general phrases of the Sherman Act were chosen intentionally, we are told by one of its framers, in order that the responsibility of determining its exact scope might be left to the courts.[1] For seven years its interpretation was uncertain. The decisions in the lower courts were conflicting, and the Supreme Court's holdings purely negative. Nor did the Trans-Missouri decision help matters much. The words "restraint of trade" still remained to be defined, and in the next thirteen years the work of definition progressed only so far as the particular cases decided were typical of classes of possible cases. The standard of public policy announced in the Standard Oil decision was the first general criterion of the scope of the act. There is nothing in subsequent decisions which indicates any tendency on the part of the Court to depart from that standard.

With restraint of trade held to be the general equivalent of monopolizing and attempting to monopolize, there is small doubt that the present interpretation of the statute is in harmony with the purposes which were in mind when it was enacted. The great industrial

[1] George F. Edmunds, "The Interstate Trust and Commerce Act of 1890," *North American Review*, vol. CXCIV, p. 801 (December, 1911). Cf. A. H. Walker, *History of the Sherman Law*, chap. I, II; and O. W. Knauth, *The Policy of the United States towards Industrial Monopoly*, chap. I.

combinations which it was hoped the statute would destroy have a more unified and compact form than they had in 1890. Whether such consolidations are contracts, combinations, or conspiracies in restraint of trade within the meaning of the common-law terms may be a matter of doubt.[1] But there is now no question but that, if their purpose is monopoly, they come within the condemnation of the Sherman Act. Nor does it appear that in any unimportant respects the statute has been weakened. There is no reason to think, for example, that price agreements and agreements to restrict output, whether of local or of general scope, are not as illegal now as they have been at any time.[2]

As a general expression of the public policy which it is supposed to embody, the Sherman Act is adequate. The difficulty is that it goes too far. In the first place, as we have seen, it is so worded that it is used as a weapon against strikes, boycotts, and other concerted efforts to interfere with the conduct of any business undertaking which ships its goods across state lines or to other countries. These things may be undesirable;

[1] Cf. the argument of Justice Holmes to the effect that under the common law contracts in restraint of trade are contracts with strangers to the contractor's business, and that combinations and conspiracies in restraint of trade have the purpose of keeping strangers to the agreements out of business. (Northern Securities Co. *v.* United States, 193 U.S. 404, 410.)

[2] The following extract from the opinion of the Circuit Court in the "bath-tub trust" case is directly in point: "Some men believe that price agreements should be sustained by the courts, unless they are shown to be against the public interest. Others hold that they may be permitted only when it is affirmatively shown that they promote the public interest. Still others say that a price agreement pure and simple is always illegal. That the Supreme Court has declared the last of the above-stated contentions to be the law is conclusive here." — United States *v.* Standard Sanitary Mfg. Co., 191 Fed. 182. See also the decision of the Supreme Court in the same case, 226 U.S. 20, and Straus and Straus *v.* American Publishers' Association, 231 U.S. 222.

some of them unquestionably are. But they are so far out of line with the other things condemned by the Sherman Act, and in most instances have so little relation to "monopolizing," that the Sherman Act ought to be so amended as to cut them out of the list of offenses which it condemns. In some cases they clearly run counter to current conceptions of public policy. In other cases this is not so clear. At any rate, if federal law is still to condemn all or some of these things, new statutes ought to be drawn. Practices like those condemned in the Eastern States Lumber Association case, mentioned above, probably fall within the jurisdiction of the Federal Trade Commission.

In the second place, the application of the Sherman Act to railroads is inconsistent with the standards of public policy embodied in the Interstate Commerce Act. We regulate railroad rates and services on the assumption that railroads are natural monopolies, that combinations or rate agreements are inevitable. But at the same time we condemn railroad combinations and rate agreements, and (as in the New Haven case) bring criminal indictments against the men responsible for such combinations.[1] From railroads we exact the observance of two mutually inconsistent standards of morality. The real evils in railroad combinations are matters of corporation finance and of the equilibrium of the equities of various sorts of bondholders and stockholders. These evils should be dealt with by statutes appropriate to the purpose, and the Sherman Act should be so amended as to be relegated to its proper field of preventable industrial "monopolizing."

[1] This, of course, has nothing to do with the questions relating to the way in which the New Haven officers and directors discharged their trusteeship for the stockholders of that road.

Finally, there comes the question of whether even within the industrial field we want to prohibit *monopoly* as well as aggressive *monopolizing*. Probably a monopoly achieved merely by the superior efficiency of a formerly competitive business unit (if such a thing were possible) would not be held to be a violation of the Sherman Act. Monopolies achieved by the active suppression of competition are illegal. What, then, is the status of a monopoly built up merely by the peaceful union or absorption of competitive units? Such combinations in the railroad field have already been condemned in the Northern Securities and Union Pacific cases. Should similar combinations in the industrial field likewise be broken up, or (since unfair methods are ruled out) should we put our trust in latent competition? On which side public policy properly lies is hard to determine. On grounds of economic principle there is nothing to fear from such combinations. But there is not much to gain from them, and in practice it would prove hard to draw a line between peaceful combination and predatory combination for monopoly's sake.

(3) Does the Sherman Act provide an efficient mechanism for achieving its own ends? That its criminal features have been ineffective is generally admitted. In many cases the criminal remedies it provides could not have been enforced except at the expense of the efficiency of the proceedings for the dissolution of the combination. Furthermore, it has been found difficult to secure a criminal conviction from a jury for an offense so general, so abstract, with so little of the common odor of crime, as "restraint of trade" or "monopolizing." It is often said that "restraint of trade" is no more indefinite than "fraud." This is very likely

so, but there is the important difference that fraud consists in, or is accompanied by, objectionable specific acts, while restraint of trade is often general and abstract. The few convictions obtained under the Sherman Act have been in cases where the restraint of trade or of commerce consisted in, or was evidenced by, specific objectionable acts. There is no reason to expect that it will ever be easy to secure convictions for restraint of trade or monopolizing in cases where the separate steps taken in the creation of the restraint are unobjectionable except as part of the general scheme.

The proceedings in equity for the dissolution of a combination have, however, proved to be increasingly effective, in the sense that there has been an increasing number of successful prosecutions. It is contended by many that the enforced dissolution of a combination means generally a mere change in form without diminution in monopoly power; that we are merely hunting the quarry from tree to tree. But in neither transportation nor industry does it clearly appear that the newer and more unified forms of consolidation would not have largely displaced the old, even if the movement had not been hastened by legislation and by decisions under the common law. Among other things tending to this end the various strategic advantages of the consolidated unit, the permanency and dependability of the newer forms of combinations, making possible the adoption of business policies based on long-time considerations, and the opportunities the single corporation, whether holding company or not, affords for the capitalization of promoters' profits may be mentioned.

Even if the Sherman Act has itself been partly responsible for the change in the form of combinations (it cannot have increased their number), it does not follow

that in the future its history is to repeat itself in this particular. It may be that, after a long chase, the quarry has finally been driven into a corner. Granting this, is it proved that the mere dissolution of industrial combinations accomplishes anything, especially in cases where the equities in the combination are made the basis of a pro rata distribution of the equities in its constituent parts?

I cannot attempt to answer so large a question. But the following conclusions appear warranted: (1) The results must vary with the nature of the business and the degree to which aggressive suppression of competition played a part in maintaining a condition of monopoly. The results must also vary with the degree to which railway rates had been shaped for the benefit of the combination. (2) Dissolution rarely comes early enough — not until the monopolistic situation (if any exists) has become more or less crystallized. This suggests the need of emphasis upon methods rather than achieved results, upon monopolizing rather than upon monopoly. (3) The operation of the statute is intermittent. Dissolution should be carefully followed up, and every step in the process of restoring normal conditions should be carefully watched. This requires administrative machinery.

II

The new anti-trust legislation comprises two statutes, the Federal Trade Commission Act, signed by the President on September 26, 1914, and the so-called Clayton Act, entitled "An act to supplement existing laws against unlawful restraints and monopolies, and for other purposes," signed on October 16, 1914. To some extent the two statutes overlap, in other respects they

are supplementary, and the Clayton Act also bears upon a variety of more or less unrelated matters which are not touched by the Trade Commission Act. Neither statute is so good a specimen of draftsmanship as the Sherman Act, while the Clayton Act, with its twenty-six sections and its heterogeneous subject-matter, is a particularly formless piece of legislation. In statute-making, defects of form are prone to be associated with defects of substance, and there is reason to fear that there is such an association in the Clayton Act.

I confess that I am not clear as to just what sort of change in the business situation the new legislation is intended to accomplish. It is clear, of course, that these statutes reflect the general public policy that is declared in the Sherman Act; that is, that competitive business conditions must be maintained and safeguarded. It is also clear that the new statutes are more detailed and specific than the Sherman Act. Moreover, additional administrative machinery is provided for the enforcement of the announced public policy. But it is not clear that the new legislation is based on any clear-cut convictions as to the particular shortcomings of the Sherman Act. This does not condemn the new legislation, for the question of the wisdom of its separate parts yet remains open. No one general purpose, but a variety of different purposes, characterizes the new laws.[1] The topics which I shall consider are (1) the exemption of labor combinations, (2) the limitations put upon the use of holding companies, (3) the prohibition of interlocking directorates, (4) the condemnation of certain trade practices, (5) the power of the Federal Trade Commis-

[1 The section of the original papers which dealt with the legislative history of the new statutes (*Journal of Political Economy*, vol. XXIII, pp. 306 ff.) is not reprinted here.]

sion over unfair methods of competition, and (6) the other powers and duties of the commission.

The sixth section of the Clayton Act, exempting labor combinations from the condemnation of the anti-trust laws, is as follows:

> The labor of a human being is not a commodity or article of commerce. Nothing contained in the anti-trust laws shall be construed to forbid the existence and operation of labor, agricultural, or horticultural organizations, instituted for the purposes of mutual help, and not having capital stock or conducted for profit, or to forbid or restrain individual members of such organizations from lawfully carrying out the legitimate objects thereof; nor shall such organizations, or the members thereof, be held or construed to be illegal combinations or conspiracies in restraint of trade, under the anti-trust laws.

The interpretation of this section may seem to be somewhat complicated by the inclusion of "agricultural or horticultural organizations." This is to be attributed to the desire of the labor leaders to secure the support of senators and representatives from agricultural States. Agricultural organizations were not actively engaged in lobbying for these provisions. Most agricultural coöperative associations are conducted for profit and so are not specifically exempted by this section. Few, if any, however, fall under the condemnation of the anti-trust laws. Agreements to maintain prices or to restrict output are not common among agriculturists. Attempts to come to such agreements, such as have been made by cotton- and tobacco-growers, have always been ineffective. The exemption of agricultural organizations has no real importance, and there

is no reason to suppose that it was intended to have importance.

The labor leaders who were active in the campaign for exemption professed to be satisfied with the section as it stands. But the opposition to it would undoubtedly have been firmer if many members of Congress had not believed that it effected no substantial change in the existing status of labor organizations under the Sherman Act. When the essential features of the present section were first decided upon, President Wilson said in an interview that it did not exempt labor organizations from the operation of the Sherman Law. Similar statements were frequently made in the debates and as frequently challenged. For a final settlement of the question we must wait, of course, for court decisions, but some points seem to stand out rather clearly.

The declaration, "The labor of a human being is not a commodity or article of commerce," [1] is little more than an empty *blague*, and the permission given to individual members of labor organizations to "lawfully carry out the legitimate objects thereof" is at once harmless and unavailing. If there is any effectiveness in the section it is in the clause, "nor shall such organizations, or the members thereof, be held or construed to be illegal combinations or conspiracies in restraint of trade, under the anti-trust laws." Now no action has ever been brought for the dissolution of a labor union under the Sherman Act. But Mr. Gompers, testifying before the House Committee on the Judiciary, said that labor combinations exist only by the tolerance of the At-

[1] The suits brought against labor unions under the Sherman Act have not charged them with monopolizing "labor." It is an amusing circumstance that the first appeal to the protection of this clause should have been made by an employers' combination — the so-called "base-ball trust."

torney-General. If Mr. Gompers's interpretation of the Sherman Act is correct, if that statute condemns unions merely because they are combinations for collective bargaining, then it must be admitted that their specific exemption in this section is a matter of some importance. But there are reasons for thinking Mr. Gompers mistaken.

In the first place, there is enough difference between the sort of "monopoly" which a labor union seeks to establish and such monopolistic industrial combinations as are clearly condemned by the Sherman Act to make it uncertain whether the courts would put them in the same category. One is in principle an inclusive, the other an exclusive, monopoly. It is true that railroad mergers, such as were condemned in the Northern Securities and Union Pacific cases, are in some respects "inclusive" combinations. But the St. Louis Terminal Railroad case is a more instructive precedent.[1] The railroad in question has a virtual monopoly of terminal facilities in St. Louis. A suit for dissolution was brought against it under the Sherman act. The Supreme Court did not grant a dissolution order, but merely directed the company so to reconstruct its organization as to provide that new companies might participate in its ownership and be given the advantages of its services on equal terms with the railroads then in control of it. The parallel is imperfect, for the Terminal Railroad of St. Louis is a natural monopoly. But the case suggests that if similar suits had been brought against labor combinations the outcome might have been that the Court would merely have insisted that admission to union membership must be granted to all applicants on fair terms.

[1] 224 U.S. 383.

In the second place, it is even doubtful whether the courts would have deemed themselves authorized by the Sherman Act to interfere in any degree either with the conditions of admission to union membership or with the ordinary trade agreements which unions attempt to enforce. Agreements among working-men to fix wages or hours or conditions of employment have only an indirect and incidental effect upon interstate trade or commerce, while the courts have consistently held that the Sherman Act covers only agreements which have a direct and primary effect upon such trade or commerce.

In the third place, there is a yet more important consideration. The Sherman Act has been brought to bear upon working-men, not because labor unions are in themselves labor monopolies, and as such are in restraint of competition, but merely because strikes and boycotts have the effect of interfering in some degree with the free flow of goods from one State to another. This, of course, is what the labor interests wanted changed. It is not clear that the new section alters the law in this respect. I cannot imagine that the courts will hold that the provision that neither labor organizations nor their members shall be "held or construed to be combinations or conspiracies in restraint of trade" covers cases of this kind. If it does, a doubt arises as to the constitutionality of the section. It is true that the Supreme Court has held that a Missouri anti-trust statute is not in violation of the Federal Constitution, despite the fact that the statute specifically exempts labor combinations.[1] But with respect to the matter in hand the question would be whether an exemption of such interference with the "free flow of commerce" as comes from the activities of labor combinations would

[1] International Harvester Co. *v.* Missouri, 234 U.S. 199.

not amount to a denial of "due process of law" to the members of such other combinations as are condemned for similarly interfering with commerce. It is not a question of the legality of the restraint of competition among working-men. The question is whether labor combinations may restrain *interstate commerce in goods* while such restraint is not permitted to other combinations. It may be that the whole line of cases in which restraint of the free movement of goods from one State to another has been condemned is based upon a twisted interpretation of the meaning of the Sherman Act. But until that statute is so amended as to apply merely to restraint of competition as distinguished from interference with commerce there seems little chance that the exemption section of the Clayton Act can alter in any important degree the present position of labor unions under the Sherman Act.

Another provision in the Clayton Act, however, may give labor combinations virtual immunity from the operations of the Sherman Act. This section (the twentieth) prohibits the granting of injunctions by federal courts in labor disputes "unless necessary to prevent irreparable injury to property or to a property right," and specifies that such injunctions shall not prohibit striking, picketing, or boycotting. This in itself does not prohibit civil suits for damages or criminal prosecutions under the Sherman Act. Probably it does not prohibit the granting of injunctions on the petition of the Government, for the general provisions of the section apply only to cases "between employers and employees." But such considerations become unimportant in view of the fact that the section concludes with the sweeping statement, "nor shall any of the acts specified in this paragraph be considered or held to be violations of any

law of the United States." This is the real exemption section. Its constitutionality is a matter about which there is some doubt, but its unconstitutionality is by no means assured. If the courts sustain it, labor unions will have been effectively freed from the restraints of the Sherman Act.

As I have already tried to show, it would have been better to give consistency and unity of meaning to the Sherman Act by making it definitely a statute against monopolizing. "Interference with the movement of commerce," whether on the part of labor unions or others, is a thing apart from the controlling purpose of the statute. It is a pity that this desirable restriction in the scope of the statute should have been brought about only indirectly and partially, and by legalizing certain specific things which had been held to interfere with the movement of commerce. If "restraint of commerce," as a thing distinct from "restraint of trade" or "monopolizing," had been stricken out of the list of offenses condemned by the Sherman Act, all of the restriction in the scope of the act which labor interests could have desired would have been accomplished without the appearance of unfair discrimination in their favor, without the suspicion of unconstitutionality, and, it may be added, without preventing the courts from restraining any direct physical interference with the actual movement of commerce.

The Clayton Act prohibits the acquisition by one corporation of any part of the capital stock of another corporation where the effect may be "to substantially lessen competition" between such corporations, to "restrain commerce," or "to tend to create a monopoly." This provision is useless, although in individual cases it may

prove to be mischievous. Ever since the Northern Securities decision, in 1904, it has been clear that the organization of holding companies for the purpose of eliminating competition or the acquisition by one corporation of stock in other corporations with the same end in view is illegal under the Sherman Act. The new statute adds nothing to the existing law except that it prohibits such acquisitions of stock in every case in which the effect may be to lessen competition between the corporations directly concerned, even though general competitive conditions may continue to exist in the industry in which such corporations are engaged. Since the new law does not represent an attempt to get rid of the holding company as a general form of business organization, the absolute condemnation of the use of this device to effect the union of two or more competing enterprises is illogical. This is a new and unwarranted departure from the general principles of the common law in regard to "reasonable restraints of trade." Very likely little use will be made of this provision. But there is a chance that it may interfere with business arrangements which would not be at variance with the general spirit and purpose of the anti-trust legislation.

Aside from the suppression of competition there are a host of evils connected with intercorporate stockholding which are not touched by the new law. The unfair treatment of minority stockholders, the general confusion in the adjustment of stockholders' and bondholders' equities, the opportunities for various fraudulent practices, and the artificial concentration of financial power, are some of these evils. Frequently these things are found at their worst in the relations between railroad companies and their subsidiaries, and here the Clayton Act specifically permits intercorporate stock-

holdings. But such matters are outside the proper field of "anti-trust" legislation. Holding companies will be necessary so long as States refuse to give certain necessary rights, such as that of holding real property or of constructing a railroad, to corporations not of their own making. Their adequate regulation must wait upon the enaction of a federal incorporation law.

The provisions relating to interlocking directorates reach a larger variety of offenses than those relating to holding companies. In fact, at this point the Clayton Act must be regarded as supplementary to the Federal Reserve and Interstate Commerce Acts as well as to the Sherman Act. The provisions affecting banks were not enacted because of any belief that there is a tendency to monopoly in the supply of ordinary commercial loans and discounts. The thing attacked was the concentration of control over such banking resources as are utilized in aiding the larger financial operations of great corporations. Such concentration of control exists, and it has not always been wisely used. But interlocking directorates among banks have in general little to do with this situation, and where they are related to it they are merely a symptom, a convenient mode of effecting coöperation.

A very common arrangement which will be affected by the new restrictions is the interlocking of the directors of two or more banks in the same city which afford different sorts of banking facilities and are hence complementary rather than competitive. Such alliances between national banks and trust companies are common. The directors who are on the boards of both banks are usually professional bankers or men who have substantial blocks of stock in each bank. There is much

to be said for such alliances, and it is hard to see what is to be gained by putting difficulties in their way.

Another provision affects common carriers and corporations or firms having dealings with them in securities or supplies or contracting with them for construction or maintenance. It is strange that so important a provision should be limited in its application to cases where officers or directors of the carrier are directly interested in the other corporation or firm. There is no obvious reason why all large transactions of the kind on the part of common carriers should not be subject to competitive bidding. In its present form the provision leaves the door open for various methods of evasion.

The most important change it may be expected to bring about is in the marketing of railroad securities. The presence of influential bankers on the directorates of American railroads has undoubtedly often helped the railroads to place their securities and to secure credit on advantageous terms. But on the other hand the banker-directors have often been able to secure substantial shares of the gains arising from various maneuvers in the field of railroad capitalization, and as a class they have been more than adequately paid for their services. At any rate, it is highly undesirable that railroads should be managed with an eye to the security market rather than to traffic and earnings. In this particular, at least, the net effect of the new regulation should be good. So also should be the effect of the provision which makes the abuse of trust on the part of officers or directors of common carriers a felony under the jurisdiction of the federal courts.

All of the foregoing provisions relating to interlocking directorates primarily affect banks or common carriers. But it is further provided that no person may be a

director of two or more industrial corporations engaged in interstate commerce if one of them has aggregate proprietorship liabilities amounting to more than $1,000,000 and if they are, "by virtue of their business and location of operation, competitors, so that the elimination of competition by agreement between them would constitute a violation of any of the provisions of any of the anti-trust laws." This is another carelessly worded provision. Except for that other provision of the Clayton Act, which, by similarly careless wording, prohibits one corporation from acquiring the stock of any competing corporation, there is nothing in the anti-trust laws which forbids the elimination of competition between two or more competitors unless either a monopoly results or the agreement is part of a general scheme to create a monopoly. And so far as this new section applies to situations where interlocking directorates might, in fact, be a means of effectively suppressing competition, it adds nothing to the Sherman Act. Combinations effected with a view to monopoly, whatever the mechanism used as the means of combination, are unquestionably illegal under present interpretations of the Sherman Act. If narrowly interpreted the new provision may prevent a large stockholder in two shoe factories, for example, from being a director of more than one of them. Aside from its possible interference with harmless arrangements of this sort, it seems to be useless.

The partial prohibition of interlocking directorates, like the similar prohibition of intercorporate stockholdings, is patchwork legislation. Among the various methods by which harmonious adjustments in corporation policies, for good or for evil, have been secured, these two (often found together) have in recent years become conspicuous. But so far as the suppression of

competition is the thing at which the new legislation is aimed, it is illogical to single out these devices for condemnation and leave such things as intercorporate leases and intercorporate sales of property in fee untouched. The use of these and other devices for the purpose of monopolizing is already prohibited by the Sherman Act. Some confusion between the relation of these devices to admitted evils in corporation finance and their bearing upon the trust situation may have been a factor in determining the attitude of Congress. American corporation law needs thoroughgoing revision, and the general status of the holding company and the qualifications of directors are among the things most in need of mending. But into this field of the internal structure of the corporation the new legislation does not go. The contribution of the provisions we have been discussing to the arsenal of weapons against the suppression of competition is small.

The trade practices condemned by the Clayton Act are (1) discrimination in the prices charged to different purchasers, and (2) tying contracts and factors' agreements. The prohibition of price discrimination leaves room for reducing prices on account of the quantity sold or a lower "cost of selling or transportation," and for "discrimination in price in the same or different communities made in good faith to meet competition." This last exception promises to create difficulties. The prohibition of price discriminations is undoubtedly aimed particularly at the practice of local price-cutting on the part of large combinations. It will be hard to draw the line between price-cutting to "meet competition" and price-cutting to suppress competition. Yet the exception is necessary. Without it those manufac-

turers and jobbers who can get their products to a given market with minimum transportation and selling costs would have an undue advantage in that market over their competitors. The tendency would be to hamper competition rather than to foster it.

The prohibition of tying contracts and factors' agreements covers leases and sales of goods (patented or unpatented) made with the understanding that the lessee or purchaser shall not use or deal in the goods of a competitor of the lessor or seller, as well as special discounts or rebates made upon such conditions. But both this prohibition and the prohibition of price discriminations apply only to cases where the effect of the prohibited arrangement may be to "substantially lessen competition or tend to create a monopoly in any line of commerce." These sections prohibit nothing not already condemned by the Sherman Act.[1] And although the new provisions are to be enforced through orders of the Federal Trade Commission, issued after hearings, the Commission has to apply to the courts for decrees making its orders effective. The general question of what sorts of arrangements may be held to lessen competition or to tend to create monopoly will be determined by the courts. But with the establishment of working precedents it may be expected that the Commission's proceedings will be prompter, simpler, and possibly more efficacious in other respects than the judicial proceedings under the Sherman Act. In some cases, moreover, it may be possible to put a stop to attempts to suppress competition before a sufficient degree of monopoly has been achieved to make it feasible to invoke the Sherman Act.

[1] Except possibly, tying contracts relating to the use of supplies or other auxiliaries in connection with a patented article.

There has been a substantial body of well-informed opinion to the effect that the prohibition of specific unfair practices, especially the local price-cut and the factors' agreement, would be an effective supplement to the Sherman Act. But this body of opinion took shape before the Supreme Court had indicated, as it did first in the Standard Oil case, that it would consider the use of such devices as prima facie evidence of an intent to violate the law against monopolizing, and before it had begun to include injunctions against the continuance of such practices in its decrees. Much of this opinion, moreover, took the form of a belief that a statute making these unfair practices criminal would prove more efficacious than the criminal provisions of the Sherman Act. But the Clayton Act does not attach criminal penalties to the violation of its prohibition of these practices. In short, it adds nothing to the substantive law on the matter. It merely provides a new procedure, and this may prove to have some advantages.

This brings us to a provision which may come to be of far-reaching importance. The Federal Trade Commission Act empowers the Commission to issue orders restraining individuals, firms, and corporations (except banks and common carriers) from using "unfair methods of competition in commerce." These orders are to be issued only after hearings, and are enforceable only through decrees of circuit courts of appeals, and are subject to appeal in precisely the same manner as the orders issued under the various provisions of the Clayton Act. Since local price discriminations, tying contracts, and factors' agreements, when they tend to suppress competition, are unquestionably "unfair methods of competition in commerce," the Trade

Commission Act overlaps the Clayton Act to this extent, and makes the special prohibition of these particular practices quite unnecessary.

But unfair competition includes other things as well. The use of bogus competitive companies, espionage of competitors' businesses, coercion in various forms, blacklists and whitelists, the securing of railroad rebates and kindred favors — these and other practices have been used in the suppression of competition and have been held to be evidence of monopolistic intent in cases brought under the Sherman Act. So far as such methods are used in connection with an endeavor to establish monopoly, they are illegal under that statute. Or, more accurately, they are held to be part and parcel of a scheme, an effort, or an achieved result, which, taken as a whole, is itself illegal. Where such practices are found to have been used in this illegal way their continuance is often enjoined in the courts' decrees.

So far as this general aspect of unfair competition is concerned, the Trade Commission Act, like the Clayton Act, adds nothing to the older statute except a new form of procedure. There is the difference, however, that it makes specific practices illegal, while the Sherman Act condemns the general purpose of the combination which utilizes such practices. This difference is probably not important. In passing upon the validity of orders of the Commission against unfair practices, in cases where the element of monopolizing is present, the courts will of necessity fall back upon precedents established in cases under the Sherman law. Of course it is possible that there will be some broadening of the notion of what constitutes unfair competition as a phase of monopolizing, but, on the other hand, not all of the practices now reached by the Sherman Act

can be made to appear as unfair methods of competition when isolated and detached from the general business schemes of which they are parts.[1] If in some cases the new procedure will make it possible to reach the monopolizing process in its early stages, much will have been gained. But it is not to be expected that all monopolistic combinations can be dealt with effectively in this way. There will still be a field for the Sherman Act.

The prohibition of unfair methods of competition in commerce is not limited, however, to cases in which the use of such methods may tend to establish a monopoly. In this respect the Trade Commission Act differs from the Clayton Act. The Trade Commission is virtually empowered to establish, through such of its orders as commend themselves to the courts, definite standards of fair competition for all business undertakings engaged in interstate and foreign commerce. This may prove to be the most important innovation in the new legislation. The full significance of this provision does not seem to have been brought out in the debates in Congress, where most of the emphasis appears to have

[1] In the "naval stores case" (Nash *v.* United States, 229 U.S. 373), in which an indictment for a conspiracy to monopolize trade by means of various unfair practices was sustained by the Supreme Court, the defense maintained that the individual acts contemplated were not in themselves illegal. Justice Holmes, delivering the opinion of the Court, said: "As to the suggestion that the matters alleged to have been contemplated would not have constituted an offense if they had been done, it is enough to say that some of them conceivably might have been adequate to accomplish the result, and that the intent alleged would convert what on their face might be no more than ordinary acts of competition or the small dishonesties of trade into a conspiracy of wider scope, as has been explained more than once. Of course this fact calls for conscience and circumspection in prosecuting officers, lest the unfounded charge of a wider purpose than the acts necessarily import convert what at most would be small local offenses into crimes under the statutes of the United States."

been put upon its bearing on the trust problem. In the debates "unfair competition" was more than once identified with "restraint of trade." But this definition of unfair competition cannot be squared with what precedents there are for the interpretation of the two phrases, although the two overlap.

In a large sense, every general statute regulating the conditions under which business is carried on helps to fix a level for competition. Labor laws, pure-food laws, and laws providing for public supervision of weights and measures, although primarily for the benefit of employees or consumers, weaken the power of unscrupulous competitors. Furthermore, the whole body of law affecting the relations between business men and those with whom they have dealings goes far to establish the rules under which competition must be conducted. But the law of unfair competition, taken in a narrower sense, bears directly upon the relations between business men as competitors; that is, upon the methods and practices used to gain trade. It establishes the lines beyond which one cannot go in the attempt to divert trade from one's competitors. Lacking statutory definition,[1] it comprises a variety of things found in different branches of the law. Combinations or conspiracies to injure or destroy a competitor's trade by the use of methods that would in themselves be legal, save for the fact of the combination or conspiracy, are, for example, to be put under this general

[1] The "unfair competition" statutes found in a number of the States relate only to price discriminations. They are in general similar to the provisions of the Clayton Act bearing upon the same matter, except that the state laws apply to all price discriminations made with the intent to injure a competitor, even where there is no purpose to establish a monopoly. These statutes are reprinted in *Laws on Trusts and Monopolies*, compiled by Nathan B. Williams for the use of the House Judiciary Committee (Washington, 1914).

head. And ordinary forms of competition may be held to be illegal if used to injure a particular undertaking and if not designed to promote the legitimate interests of the persons employing such methods.[1] These common-law doctrines have a wider application than that suppression of competition or "monopolizing" which is condemned by the Sherman Act. Not the prevention of monopoly, but the protection of the individual business undertaking is their controlling purpose. The kind of competition which they condemn need not have for its objective the establishment of monopoly; it is sufficient if its direct and primary purpose is to injure or destroy.

Then there is the very different sort of competition which is condemned, not because of its general purpose, but because of the methods which it employs. Inducing a competitor's customers to break their contracts with him is a case in point. But more important are such things as libelous statements and fraudulent misrepresentations. By far the largest number of cases in the general field of the law of unfair competition have to do with fraud, and more especially with methods calculated to enable the offender to trade upon the established good will of a competitive undertaking. The use of trade-marks, brands, firm names, packages of a particular form or appearance, etc., for the purpose of misleading the consumer with respect to the identity of the firm or the origin of the goods is the most common offense of this sort.[2]

I have made this cursory and incomplete review of the general meaning of "unfair methods of competi-

[1] For a general review of the law on these points see Bruce Wyman, *The Control of the Market*, chaps. II, III, V.

[2] For an excellent popular account of these matters see E. S. Rogers, *Good Will, Trade Marks, and Unfair Trading* (Chicago, 1914).

tion" merely to suggest the extent to which the Trade Commission Act seems to pass beyond what has been the accustomed province of anti-trust legislation. But it is none the worse on that account. Aside from the additional protection which the Commission's power in these matters will give to the rights of business men and of consumers, anything that will raise the general level of competitive standards must be welcomed. Moreover, the fact that the statute's condemnation of unfair competition is not made to depend on a proved purpose to create a monopoly will make it a more effective weapon against monopoly itself in its early stages. In this way it may prove a valuable supplement to the Sherman Act.[1]

It should not be imagined that the Commission will be able to create offhand a general code of definitions of unfair practices, or that it can make substantial additions to the existing law on the subject. Its orders are subject to review by the courts on all matters of law, and the courts will, of course, define "unfair methods of competition" in the light of existing judicial precedents. The law will grow, as other laws grow,

[1] While only a few countries have taken a position as definitely opposed to industrial combination as the United States has in the Sherman Act, a number of countries have preceded us in adopting statutory declarations against unfair competition. A thing very commonly prohibited is the factors' agreement in its various forms. See, for example, the statutes of Australia, New Zealand, and Canada, reprinted in Williams's *Laws on Trusts and Monopolies*. The recent comprehensive statutes of Germany (1909) and Denmark (1912) relating to unfair competition are reprinted in Hearings before the House Committee on the Judiciary, 63d Congress, 2d session, on Trust Legislation, pp. 1491, 1495. The most interesting departure in both statutes is the careful limitation and regulation of price-cutting "sales." [For a much more complete international survey see *Trust Laws and Unfair Competition* (Washington, 1915), written by Francis Walker and other members of the staff of the Bureau of Corporations, but published under the name of Joseph E. Davies, who was Commissioner of Corporations at the time.]

only as individual cases with new characteristics are brought under it. But as to the wisdom of the general policy it embodies there can be, I imagine, no difference of opinion.[1]

The duties of the new Federal Trade Commission are not limited to issuing orders against the use of unfair methods of competition. Composed of five members holding office for seven years each, it falls heir to the corps of employees and the unfinished tasks of the Bureau of Corporations.

That Bureau, established by Act of Congress in 1903 in the Department of Commerce and Labor, was authorized "under the direction and control of the Secretary of Commerce and Labor" to investigate the transactions of any corporation or combination engaged in interstate commerce, with the exception of common carriers. Its powers in the way of compelling testimony and the production of books and papers were identical with those of the Interstate Commerce Commission. It was also authorized to gather and publish "useful information" concerning corporations engaged in interstate commerce. The primary purpose of its special investigations of particular corporations and combinations, as stated in the statute creating it, was "to gather such information and data as will enable the President of the United States to make recommendations to Congress for legislation for the regulation of commerce," and the President might decide what part

[1] It is possible that either the Commission or the courts will refuse to give so broad a meaning as I have suggested to the phrase "unfair methods of competition." But it can hardly be expected that it will not be held to cover much more than the types of unfair competition which have figured in cases under the Sherman Act. [For information with respect to subsequent developments see the *Annual Reports of the Federal Trade Commission*, together with *Public Regulation of Competitive Practices* (National Industrial Conference Board), 1925.]

of its information should be made public. Established at the behest of President Roosevelt, it was to be an arm of the executive rather than of the legislative or judicial branch of the Government. In addition to performing the duties with which it was legally charged, it has coöperated with the Department of Justice in various ways, and some of its investigations have been undertaken at the special authorization of Congress.

The new Commission is not a bureau of any department, but has an independent status, like that of the Interstate Commerce Commission. Some of its work, however, is as a virtual auxiliary of the Department of Justice. At the request of the Attorney-General it is to investigate any corporation alleged to be violating the anti-trust laws, and to make recommendations for the readjustment of its business. This is a wise provision for informal adjustments like those which are so important a part of the work of the Interstate Commerce Commission. In suits in equity brought under the anti-trust acts the Commission may be asked by the Court to prepare an appropriate form of decree, which is, of course, subject to rejection or change by the Court. The importance of this provision is in its bearing upon the outcome of dissolution proceedings under the Sherman Act. Since the American Tobacco case, it has been recognized that the drafting of a wise plan of reorganization for an offending combination may be an exceedingly difficult matter, requiring not only care and judgment, but also a large amount of technical information about the general conditions of the industry affected.

Like the former Bureau of Corporations, the Commission may investigate and report upon the affairs of industrial and trading corporations engaged in inter-

state commerce, and it may be asked by the President or either House of Congress to report upon any alleged violation of the anti-trust acts. It has the further power to require annual or special reports from interstate corporations in such form and relating to such matters as it may prescribe. The information it obtains may be made public at its own discretion, "except trade secrets and names of customers." One effect of this provision, it may be hoped, will be to give students of corporation problems a more adequate body of authentic and apposite information than they have yet had.[1]

It is hazardous to pass a general judgment upon legislation which reaches into fields so new as do these two statutes. But unless the foregoing analysis is altogether mistaken the two cannot be joined for either praise or condemnation. The anti-trust sections of the Clayton Act are bungling and generally futile. There is a chance that, at the worst, they may make enough trouble to delay the enaction of the badly needed federal statute dealing thoroughly and systematically with the promotion, organization, and management of corporations engaged in interstate commerce. At best, so far as I can see, they will be ineffective. The Federal Trade Commission Act is a statute of a different type. It introduces a method of dealing with the admitted evils of unfair competition modeled upon what has proved a successful method of dealing with railroad discriminations. The new machinery it furnishes for the enforcement of the Sherman Act is of a kind which

[1 This hope has thus far (1927) been disappointed. Accounts of the work accomplished by the Commission are given by Gerald C. Henderson, *The Federal Trade Commission* (New Haven, 1924); Myron W. Watkins, "The Federal Trade Commission," *Quarterly Journal of Economics*, vol. XLI, no. 1 (November, 1926).]

experience has shown to be needed. It makes a beginning, at least, in providing for adequate federal statistics of industrial corporations. It should prove a serviceable addition to the Nation's industrial code.

X

SOME LIMITATIONS OF THE VALUE CONCEPT[1]

EVER since the publication of *The Wealth of Nations* the notion of exchange value has remained the most fundamental and possibly the most consistently defined concept of economic science. Bickerings as to whether "power in exchange," "quantity received in exchange," or "ratio of exchange" best expresses this concept, as to whether exchange value is a variable attribute of a commodity or merely a disembodied ratio, have not been taken as seriously affecting its precision. Possibly these differences, if consistently followed out, might have led to various differences in doctrine. But historically, if not logically, the differences alluded to have been verbal. I do not propose to suggest any revision of an elementary concept which has worn well in actual service. I shall venture merely to suggest what some of its limitations are.

The notion of exchange value has a superficial resemblance to some of the abstractions of physical science, such as mass, extension, energy, and the like. Analogies based on this resemblance have sometimes been stressed by economists, especially in discussions of the "measure of value." Like value, these physical abstractions figure as properties or qualities of things, detached by analysis from the synthetic ideas given directly by experience. Abstraction makes it possible for science to find order and simplicity in an apparently

[1] Reprinted, with minor changes, from the *Quarterly Journal of Economics*, vol. xxv, no. 3 (May, 1911).]

complex universe. Furthermore, abstractions enter into all thinking which deals with general principles. These commonplaces of logic are rehearsed lest the trend of the present discussion should be taken to involve a necessarily futile argument against the use of abstractions in economic theory.

The concrete facts are the exchanges of goods and services for money and money substitutes. Prices, not values, are the primary elements of the situation with which the economic theorist has to deal. And as it happens, economic theory does concern itself very largely with problems of price. The traditional theory of distribution deals with the forces determining the prices paid for the services of productive factors. The theory of "normal value" is essentially a theory of price tendencies. The theory of market value commonly appears as a demand and supply theory of prices.

But the real significance attached to money prices has been less than one would infer from their apparent prominence. Price has figured in large part as a convenient and workable substitute for the supposedly more general, though more cumbersome, notion of exchange value. So far as the theory of exchange has been couched in terms of price, it has been felt that the procedure involved some sacrifice of scientific rigor, and was to be justified only because it simplified the task of exposition.[1] That "price is value expressed in terms

[1] Even Marshall, whose theory is cast more consistently in terms of price than that of any other writer since Cournot, explains (*Principles of Economics*, 5th ed., p. 62) that in his treatise "the price of anything will be taken as representative of its exchange value relatively to things in general, or in other words as representative of its general purchasing power." A more explicit statement is Pareto's (*Manuel d'économie politique*, p. 209): "The general notion of the price of one commodity in terms of another is useful in economic science because it eliminates money. In practice the prices of all commodities are stated in terms of

of money" is the common formula by which the transition from one concept to the other is effected. Ignore changes in the general purchasing power of money and prices become accurate measures of exchange values and hence convenient substitutes for them.

In short, value is generally thought of as logically antecedent to price. Value is the primary, price the derivative concept. This subordination of price to value puts aside the fact that exchange values emerge only from the actual process of exchange (and there emerge as prices) as of minor import for the purposes of pure theory. The only excuse for this procedure is that value is the more general of the two concepts. But this greater degree of generality is purchased at the expense of precision and reality.

In most of the definitions exchange value falls into one or the other of two categories. In the one it appears as the general purchasing power or general ratio of exchange of a commodity, a notion which is vague and abstract in the highest degree. The makers of index numbers have had to explain repeatedly that they are dealing only with the *variations* of prices; that a measure of the general purchasing power of money is virtually meaningless.[1]

In the other class of definitions value appears, not as general purchasing power, but as any one of a miscellany of purchasing powers. In this sense a commodity is

one of them, which is called money, so that it is difficult to avoid speaking of price in this sense when one is discussing concrete phenomena. Even in theory it is very useful to introduce this notion at the first. This, of course, anticipates the theory of money, which ought to come after the general theory of economic equilibrium, but there is no great harm in this, especially if the increased lucidity in exposition which the use of this concept gives is taken into account."

[1 This does not mean that it is impossible to measure *changes* of the general purchasing power of money. See below, p. 261.]

said to have not one but many values, corresponding to its ratios of exchange with all other commodities. Price, or money value, appears merely as one species of a large genus. It is fairly clear that this notion of value is really derived, by analogy, from the notion of price. Given the ratio at which each of two commodities exchanges for money and it is a matter of simple arithmetic to determine the "value" of each in terms of the other. Clearly here is something more tangible than the notion of "general purchasing power." But this other notion is not free from difficulties of its own.

It leads to the elimination of money as an essential part of the mechanism of the market. Values are pictured as determined as though under a régime of pure barter. Money is brought in at the end (for exposition's sake) as a register or common denominator of the values reached.

Various objections to a view from which so many of the dominant facts of the actual market have been abstracted suggest themselves. That a state of pure barter is, even historically speaking, probably a sheer work of the imagination, is perhaps not a relevant objection if it can be shown that projecting an idealized scheme of barter into the framework of our money economy makes for simpler analysis. For a similar reason I do not attach much weight to the obvious objection that the complexity of the present system of division of labor and exchange renders a general system of barter altogether unthinkable. Of more cogency is W. C. Mitchell's suggestion [1] that the money concept has itself been an active factor in giving purpose, system, and rationality to economic activity. Modern business

[1] In his paper, "The Rationality of Economic Activity," *Journal of Political Economy*, vol. XVIII, pp. 208 and ff.

is conducted by men who have learned to think in terms of money, and the price-making process is largely in their hands. But this objection presupposes a point of view quite different from that of current economic theory. Within the conventional limits which economic theory has set for itself can there be found reason for questioning our right to eliminate money from a general view of the valuation process? I believe that such ground of dissent exists.

The theory of exchange value is essentially a theory of the equilibrium of demand and supply. The values which would be determined if equilibrium could be achieved are consistent one with another. That is, in order to ascertain the value of A in terms of B, it would be sufficient to know the values of A and of B in terms of C. Now this fundamental postulate of a tendency toward a static equilibrium is inconsistent with the other postulate of a state of pure barter. In pure barter there is no efficient tendency toward a definite equilibrium. So far as the ratio of exchange between any two commodities is concerned an equilibrium point might be reached, although the market would have to be fairly large. But it would be quite unreasonable to expect that the $n(n-1)$ ratios of exchange thus established should be mutually consistent.[1]

The fact is that in treating exchange values as money prices in the analysis of supply and demand we are really doing more than availing ourselves of a convenient method of exposition. We are using a necessary and integral part of the analysis. The lucidity which the premising of a general medium of exchange adds to

[1 Cf. L. Walras, *Éléments d'économie politique pure*, 4th ed. pp. 118, 119. When he introduces arbitrage transactions Walras takes the first (and most important) step away from the assumption that exchanges are effected without the use of money.]

economic analysis (as in the theory of supply and demand *at a price*) is only a reflection of the precision and determinateness which the use of money gives to the actual operations of the market.

The subordination of price to value may be a heritage of the interest which Adam Smith and his immediate followers took in the problem of "the real measure of value." [1] Both the theory of national wealth and the theory of distribution seemed to demand a better measure of wealth than was afforded by money prices, with their continual fluctuations. From this source two divergent streams of theory have flowed. On the one hand we still seek for an "ultimate standard of value." This is a practical problem, discussed in the literature dealing with index numbers and with the standard of deferred payments. It has no direct relation to the problem of economic equilibrium. On the other hand we have continued to busy ourselves with the problem of the real measure of value in the sense of the *exact* measure of value. Here the marginal utility concept now holds the field. But there is nothing in the new analysis that prevents a frank recognition of the dominance of the rôle played by money prices in the system of economic equilibrium. From utility, up through marginal utility, subjective value, and exchange value, to price, is a long and slippery road. Marginal utility, like price, is a relatively simple notion, close to the concrete facts of experience. Exchange value is a looser and thinner abstraction.

Frank recognition of the fact that the notion of exchange value is a derivative of the phenomena of price would call for no substantial modification of the

[1] See on this subject, T. S. Adams, "Index Numbers and the Standard of Value," *Journal of Political Economy*, vol. x, especially pp. 13 and 14.

theory of exchange. What now passes as a concession to the exigencies of exposition would stand forth without apology as a prime factor in the situation. This would be a wholesome change of emphasis.[1] For there are indications that we have long since passed the point of diminishing returns, so far as added refinements in the general theory of static equilibrium are concerned. What is needed is an analysis of the actual mechanism of the price-making process. There should be no room for such crudities as even an implied determination of prices by the comparison of the "values of commodities" and the (independently determined) "value of money."

When we leave the pure theory of exchange and pass to that general group of problems and theories in which value figures as a measure of the existing stock of wealth or of any part of that stock we encounter difficulties of a different sort. Here economists have generally been content to speak of value in the sense of money value. This may be due to the practical difficulties in the way of any other procedure, or to a realization that the summation of "exchange values" is akin in principle to an attempt to determine the weight of the solar system. Yet even in this sense value is not equivalent to price. Prices emerge, as concrete facts,

[1] In 1909 the *Railway World* obtained the opinions of a large number of economists on the question whether there could be said to be a distinction between "real prices" and "nominal or money prices," corresponding to the common distinction between "real wages" and "nominal wages." The majority of the answers were in the affirmative. C. W. Mixter went so far as to imply that money prices were "apparent" and "real prices" were actual. On the point at issue I find myself in accord with H. J. Davenport, that "to talk of real as against nominal prices is terminological nonsense," and with D. R. Dewey, who said, "The question is absurd. There is but one kind of prices."

only in the process of exchange, and only in the process of exchange does money actually "measure value." Value, as applied to a stock of goods, is nothing more or less than *imputed price*.[1]

The grounds and purposes of the imputation vary. It may be a merchant's appraisal of his stock on hand, the assessment of property for taxation, or its "valuation" for purposes of public control; it may arise from a statistician's efforts to reach an estimate of national wealth; finally, it may be implied in such economic theories as those in which "capital" is made to consist of all wealth, measured in terms of money value. Leaving aside the cases in which the imputation is a judgment of "what price ought to be,"[2] the general method of imputation may be said to consist of establishing hypothetical prices for all the units of a stock of goods on the basis of the current prices received and paid for similar units of similar goods. In the case of non-reproducible goods, and to a greater or less extent in the case of all non-standardized goods, the process of imputation involves varying degrees of estimate and guesswork. Moreover, as will be indicated presently, much depends upon assuming that "similar units" as well as "similar goods" are to be dealt with.

It may be objected that in fact other methods of valuation than the imputation of current prices to a stock are used. "Physical valuation," for purposes of public control might, for example, be cited. But physical valuation is not an attempt to get at the price of a business undertaking considered as a unit. When it

[1] Irving Fisher's definition of value restricts it to this particular usage. "The value of goods is the product of their quantity multiplied by their price." See *The Nature of Capital and Income*, p. 336.

[2] Corresponding to the definition of value which President Hadley prefers. See his *Economics*, p. 92.

deals with "reproduction costs" it is an imputation of current prices to the separate parts of the material equipment of a business unit. When it deals with "the original cost of production" physical valuation is simply a measure of the investment of capital. Or, again, it might be urged that the valuation of income-yielding goods is a capitalization of the income yielded. This point has a direct bearing on the assumptions of certain latter day economic theories.

That the method of capitalization is used as a method of imputation cannot be denied. But what the method affords is not a determination of exchange value, but merely an estimate. Among the things bought and sold in the market are securities and commodities that have more or less definite potencies in the way of conveying money incomes to their owners. Market forces establish prices for such commodities and securities, and hence establish ratios between incomes of given amounts and of given degrees of certainty, spontaneity, and futurity, and the prices of the income bearers. In the absence of evidence more directly to the point, the ratios or rates of capitalization thus established for the incomes from certain kinds of goods and securities may be used as the means of imputing price to similar goods and securities, not currently priced in the market, but with known or estimated income-yielding power.

The possible objection that the "forces of the market" are themselves in large part the resultants of the ratios at which individuals equate future incomes to present values is not to the point. For we are dealing with market values, which, so far as each individual is concerned, must be held to be objectively determined and objectively measurable. Considerations essentially similar to those outlined above hold, *mutatis mutandis*,

with respect to the pseudo-capitalization of "psychic income." [1]

The value of a stock, then, as an economic concept, may be taken to be a derivative of the fact that particular units of the stock are exchanged or are estimated to be *exchangeable* at certain prices. The value of a homogeneous stock is got by multiplying the number of units in the stock by the price per unit. For the most part economists have not thought it necessary to make a distinction between the price of the units exchanged and the exchange value, or imputed price, of the units not exchanged. But Cournot [2] thought it worth while to call attention to the simple and elementary fact that the maintenance of the level of current prices is dependent on not overcrowding the market. Not all the units of any considerable stock, such as the stock of wheat, could be sold at one time at current prices, nor on the other hand could buyers of any good substantially increase their purchases except at increased prices.

The notion of the value of a stock seems to hinge on the fact that presumably any one unit of the stock may be added to the amount sold, or to the amount bought, without substantially affecting the price. The imputed price seems to hold rigidly for the particular units of a stock, taken one at a time. If it were not that it would put an additional burden on an already hard-

[1] The foregoing discussion is, I think, in essential harmony with the thesis maintained by H. J. Davenport, in his paper on "Capitalization and Market Value," *Yale Review*, vol. XIX, p. 132 (Aug., 1910). [Cf. also Professor Davenport's *Economics of Enterprise*, chap. XV.]

[2] "Under this conception [a sum of exchangeable values] wealth has doubtless only an abstract existence; for, strictly speaking, of all the things on which we set a price, or to which we attach a value in exchange, there are none always exchangeable at will for any other commodity of equal price or value." — *Mathematical Principles of the Theory of Wealth* (Eng. transl.), p. 9.

worked adjective, it might be insisted that exchange value, after all, is only *marginal value*. For all except the marginal units the imputed price is purely hypothetical, subject only to the limits set by the existing potential demand and existing potential supply.[1] As Jevons suggested,[2] the theory of demand and supply is properly a theory of *rate of demand and rate of supply*.

A stream of goods flows through the market from sellers to buyers, and at the point of exchange this flow is equated (in terms of price) to the stream of money flowing in the opposite direction. The vague outlines of a dynamic theory of price are easily imaginable. Such a theory might analyze the forces controlling the volumes and rates of flow of particular kinds of commodities, and the volumes and rates of flow of the parts of the money stream to which these are equated in the market. Such a theory would not lead to conclusions substantially different from those reached by the analysis of the forces tending to static equilibrium, and it would be decidedly more cumbersome. But the use of the static method has tended to falsify our view of the facts in some particulars.

Seizing a moment when the two streams are running smoothly and steadily (corresponding to the condition of static equilibrium), we imagine them, in effect, to be suddenly congealed. Then, with this tactical advantage, we devote ourselves to a painstaking analysis of the proximate factors determining the prices of the goods which happened to be thus arrested at the very

[1 Professor Davenport argues (*Economics of Enterprise*, p. 385, note) that I neglect holders' "demands" and their "reservation prices." Not so. I am merely emphasizing the truism that holders' reservation prices are above the market, just as the prices at which buyers will increase their purchases are below the market.]

[2 *Theory of Political Economy*, 3d ed., p. 64.

moment when they were passing through the narrow channel of exchange. But to be satisfied with this achievement would be to fall short of the opportunity for system making. So we examine the upper reaches of the congealed stream of goods, imputing value to everything we find, on the basis of the price units discovered at the point of exchange. Going still farther back we subject to the same Midas-like touch the upper reservoirs of goods that are usually drawn upon only when the stream is running dry. And finally, by a supreme *tour de force*, we convert into value units those outlying pools of intimate personal belongings, not ordinarily appraised in terms of money value, even by those who prize them most, and from which normally only a thin rivulet trickles to join the stream of goods passing through the market place.[1]

In short, for system's sake, the whole material equipment of human living is recast in molds fashioned after the notions of catallactics. This view of things is implicit in a large part of the body of systematic economics. But I wish frankly to say that in my opinion the symmetry and logical completeness of the systems of Professor Fetter and Professor Fisher are due to the fact that they have gone farthest on this road.

If the value of a stock is a matter of imputed price, and if this is really an accurate measure only for the marginal units of a stock, the size of the unit becomes a matter of some importance. This suggestion should not be confused with the objections urged by Mr. Hobson and others against the theories of marginal

[1 This means that the economist has gone further in "objectifying" value than the facts warrant. Cf. E. E. Hoyt, *Primitive Trade*, chap. v.]

productivity and of marginal utility on the fallacious ground that the results reached depend upon the size of the units chosen. For these theories deal with changing ratios (of utility to quantity or of product to labor) and in whatever form they are stated there is implied in them the mathematical notion of a limiting ratio. There is no implication of this sort in the ordinary notion of the value of a stock of goods. Here we have to deal with the imputation of market prices to concrete units of definite size.

An example taken from actual experience may serve to make the matter clear. A forty-acre tract of land on the outskirts of a small village was assessed for taxation at $8000, or $200 per acre. Half of one acre was subsequently sold as a building lot, and assessed at $400, or $800 per acre. As the building lot was not taken from the best part of the tract, its owner made complaint of overassessment. A little investigation showed that though several more building lots of the same size could be sold at or near a price equivalent to $800 per acre, the tract as a whole could not be sold for much more than $200 per acre. Any one acre, taken by itself, was worth $800; yet to have multiplied this amount by the number of acres would have given the absurdly high total value of $36,000. What, by the method of price imputation, was the total value of the tract?

This illustration involves more than the principles implied in the imputation process, i.e., that the market must not be overcrowded and that the rate of supply is supposed to be normal. The fact seems to be that the price of building lots and the price of the tract as a unit were affected by and adjusted to entirely different fields of demand. The example is probably an extreme one. But that the total value of a tract of land depends

in part upon the way in which it is subdivided is a familiar fact of real estate operations.

The field of corporation finance furnishes similar problems. The sum of the value of the disintegrated parts of a corporation's material equipment together with the value of its franchise, if this can be sold separately, is one thing; the price which the corporation's assets as a whole would bring under the hammer is another thing; and the total value of the outstanding securities of a corporation may be yet another thing. Here again the essential thing in the situation is the fact that the equities in the corporation's property may be subdivided in different ways, and that the field of demand varies with the method of subdivision.

Take, for example, the problem of the value of a corporation's securities. Here, as in the case of some railroad corporations, the bulk of the stocks and bonds may be securely held for purposes of investment and control. The value imputed to these is derived from the prices of the relatively small part of the securities which happen to figure actively on the stock exchange. Disregarding the element of control, and assuming that the rate of supply is not under any circumstances likely to be extremely high, imputation may be said to give fairly precise results in such cases. For such securities are distinctly marketable goods, and their market is highly organized and elastic.

But the fact remains that the total value of the equities in a corporation depends on the way in which they are subdivided. Bonds of small denominations find a market not open to bonds of larger denominations. Moreover, in a complexly capitalized corporation, the classification of securities is such that it offers to buyers of incomes a carefully graded assortment of risks.

This maximizes the selling or capital value of a corporation's income-earning power. Different levels of demand are tapped and the result is better than if the curve of diminishing buyers' prices for any one level were followed too far. Preferred stocks may sell readily when bonds are a drug in the market. That changes in total value are affected by a skillfully conducted railroad reorganization (even when no new capital is invested and net earning power is not increased) is well known.

All these considerations may have some bearing on practical problems of assessment and of public valuation. Possibly they deserve consideration in connection with any theory which deals with the notion of "wealth as a sum of exchangeable values," although I am disposed to regard them as relatively less significant than the more general views suggested earlier in this paper.

XI

JEVONS'S "THEORY OF POLITICAL ECONOMY" [1]

NONE of the various filial services which Professor H. S. Jevons has performed in the way of rounding out his father's work and making it more accessible will be received more gratefully by economists than the new edition of his most important book on economics.[2] Except for a few corrections and explanatory notes the body of the book is a reprint of the third edition, which has been out of print for some years. The promised extension of the bibliography of mathematical economics does not carry it beyond the year (1879) to which it had been brought in the third edition by Mrs. Jevons. Some new titles have been added, but the list as it stands is far from complete. Some interesting and previously unprinted notes by W. S. Jevons on a few of the writings of his predecessors and contemporaries have been added as appendices.

It happens that it was just fifty years ago that Jevons, then twenty-seven years old, sent to the British Association meeting at Cambridge a short paper entitled, "Notice of a General Mathematical Theory of Economy." This was accompanied by another paper on "The Study of Periodic Commercial Fluctuations." Jevons's innovations in theory do not seem to have at-

[1] Reprinted, with minor changes, from *The American Economic Review*, vol. II, no. 3 (September, 1912).]

[2] *The Theory of Political Economy*, by W. Stanley Jevons, *with Notes and an Extension of the Bibliography of Mathematical Economic Writings* by H. Stanley Jevons. (Fourth edition. London: Macmillan and Company, Ltd., 1911, pp. lxiv, 339.)

tracted much attention at the time of this, their first public presentation. He wrote to his brother that he was informed by the secretary that both papers "were read before the F Section, and the second [the statistical study] was approved of."[1] Jevons had become interested in political economy in the course of the extensive reading which filled most of his spare hours during the years from 1854 to 1859, while he was serving as Assayer of the Mint at Sydney, Australia. His conviction, then formed, that the field of "social science" had been unsatisfactorily tilled by previous writers and that it was a field in which he might hope to accomplish work of importance, were among the reasons which led him to resign a fairly lucrative position and to return to England and an uncertain future. During the four years immediately following, Jevons completed his work for the B.A. and M.A. degrees at University College, London, where political economy continued to claim a considerable share of his attention. Two letters written to his brother in 1860 are worth quoting:[2]

> During the last session I have worked a good deal at political economy; in the last few months I have fortunately struck out what I have no doubt is *the true Theory of Economy*, so thoroughgoing and consistent, that I cannot now read other books on the subject without indignation. While the theory is entirely mathematical in principle, I show, at the same time, how the data of calculation are so complicated as to be for the present hopeless. Nevertheless, I obtain from the mathematical principles all the chief laws at which political economists have previously arrived, only arranged in a series of definitions, axioms, and theories almost as rigorous and connected as if they were so many geometrical problems. One of the most important axioms is, that as the quantity of any

[1] *Letters and Journals of W. Stanley Jevons*, edited by his wife, p. 169.
[2] *Letters and Journals*, pp. 151, 155.

commodity, for instance, plain food, which a man has to consume, increases, so the utility or benefit derived from the last portion used decreases in degree. The decrease of enjoyment between the beginning and end of the meal may be taken as an example. And I assume that on an average, the *ratio of utility* is some continuous mathematical function of the quantity of commodity. This law of utility has, in fact, always been assumed by political economists under the more complex form and name of the Law of Supply and Demand. But once fairly stated in its simple form, it opens up the whole of the subject. Most of the conclusions are, of course, the old ones stated in a consistent form; but my definition of capital and law of the interest of capital are, as far as I have seen, quite new.

. . . I expect every success from my theory of political economy, which seems to develop itself with that facility which is a proof of its soundness. It assumes the form of a complicated mathematical problem, from which all the common laws with due limitations flow. Independently, however, of the mathematical form, it has led me to a new view of the action of *capital*, which affords a determining principle for *interest*, profits of trade, wages; and I now perceive how the want of knowledge of this determining principle throws the more complicated discussions of economists into confusion. The common law is that demand and supply of labour and capital determine the division between wages and profits. But I shall show that the whole capital employed can only be paid for at the same rate as the *last portion* added; hence it is the increase of produce or advantage, which this last addition gives, that determines the interest of the whole.

These letters are significant, not so much because they have some bearing upon the more or less futile questions relative to Jevons's priority in the formulation of certain theories or the use of a particular method, as because they foreshadow so completely the tenets of *The Theory of Political Economy* of 1871. The book con-

tains very little of importance beyond the matters mentioned in the letters of 1860. The printed abstract of the British Association paper of 1862,[1] in fourteen separate propositions, makes clearer the hedonistic basis of the new theory, explains the nature of the proposed mathematical equations, states the new definition of capital, and adds a tentative theory of wages. During the next few years Jevons gave his spare time to various studies in statistics and in logic and to his book on *The Coal Question*. When, in 1866, his paper on "A General Mathematical Theory of Political Economy" was first published *in extenso*,[2] it had received no revision. Logical studies, including some preliminary work on his *Principles of Science*, continued to absorb his attention until the winter of 1870–71, when, Mrs. Jevons tells us,[3] "he returned with renewed interest to political economy and devoted himself entirely to the writing of *The Theory of Political Economy*. This work was of such absorbing interest to him that he made rapid progress with it." Some correspondence with Fleeming Jenkin relative to the use of the mathematical method in economics and the publication, in 1870, of Jenkin's "Graphic Representation of the Laws of Supply and Demand" in Grant's *Recess Studies* seem to have led Jevons to hasten this completer presentation of his own theories.[4]

Despite Jevons's growing reputation, the book was slow to make a place for itself. In 1874 Jevons wrote to M. d'Aulnis de Bourouill of the University of Ley-

[1] *Report of the British Association for the Advancement of Science*, 1862, Notices and Abstracts, p. 158.

[2] *Journal of the Statistical Society*, vol. xxix, p. 283. Reprinted as Appendix III in the present edition of the *Theory*, pp. 303–14.

[3] *Letters and Journals*, p. 251.

[4] H. S. Jevons's preface to the present edition of the *Theory*, p. lvii.

den, "What I have written on the subject of mathematical economics has received little or no attention in England, and by those who have noticed it the theory has been generally rejected, or even ridiculed." [1] The *Saturday Review* had printed a review of the *Theory* by Cairnes, who did not let his confessed inability to understand much of the argument prevent him from handling it severely. Marshall had contributed to the *Academy* an able and discriminating review, which, although fair, was in no manner enthusiastic. The only English writer of importance who definitely espoused the cause of the mathematical method in general and Jevons's theories in particular, was G. H. Darwin, who, in reviewing Cairnes's *Leading Principles* in the *Fortnightly Review* in 1875 defended Jevons's *Theory* against Cairnes's criticisms. It seems, however, that from the first the book had attracted somewhat more attention on the Continent than in England. Appreciative letters from such men as Walras, Bodio, and d'Aulnis de Bourouill and the ensuing correspondence with them must have done much to support Jevons's confidence in the soundness and importance of his own work and to quicken his desire to extend his analysis to other parts of the field of economic theory. Even his discovery that he was not the only one nor even the first in the field — that the very fundamentals of his theory had been anticipated in the forgotten writings of Gossen and Dupuit; that Walras, working independently, but almost contemporaneously with himself, had utilized methods and reached conclusions which were in many respects like his own; that Cournot had applied the mathematical method to a somewhat different class of problems with unequaled power and fluency — whetted rather than dulled his enthusiasm.

[1] *Letters and Journals*, p. 309.

It was not long, however, before the *Theory* began to gain more serious consideration in England than it had received at first. Jevons's standing as a logician had been firmly established by the publication of *The Principles of Science*, and his continued output of brilliant studies on a wide variety of topics reached an appreciative and growing audience. In short, he had become a man to be reckoned with. The revolt against the political economy of the orthodox line, headed by Thornton, and continued by Cliffe Leslie and Ingram, led some few persons, at least, to look for a new economic gospel. And the growing influence of Marshall's teaching at Cambridge was not without its effect. In 1875 Jevons wrote to Walras:[1]

> I think that a considerable change of opinion is taking place in England. Various correspondents express their acquiescence, and some of the professors are beginning to bring the theory before their students. When I was in Cambridge two months ago I found that the subject was much better understood there than I had supposed, and I have little doubt about its gaining ground ultimately. . . . I have no doubt whatever about the ultimate success of our efforts, but it will take some fighting; the disciples of J. S. Mill being bitterly opposed to any innovation upon his doctrine.

The second edition of the *Theory* (the last edition during Jevons's lifetime) appeared in 1879. The most important additions besides the bibliography of mathematical economics and the lengthy and interesting preface, are the discussions of the "dimensions of economic quantities." Here Jevons illustrates to perfection his

[1] *Letters and Journals*, p. 332. In 1881 he wrote to the same correspondent, "I am glad to say I think the mathematical view of economics is making much progress in England, and is fully recognized by those competent to judge." — *Ibid.*, p. 431.

extraordinary power of raising new questions without solving them. In some respects his treatment of the "dimensions of economic quantities" is perfunctory; in other respects it is suggestive, but incomplete and inaccurate.[1] But for the most part the second edition follows closely the lines of the first, and the first is little more than a working-out of the theses announced in 1862 and foreshadowed in the letters of 1860.

These facts, fairly well known for the most part, in the history of a book now become a classic, suggest in themselves the strength and the limitation of the work. In the preface to the second edition Jevons stated that the book "was never put forward as containing a systematic view of economics. It treats only of the theory and is but an elementary sketch of elementary principles. The working-out of a complete system based on these lines must be a matter of time and labor, and I know not when, if ever, I shall be able to attempt it." [2] Even more accurate is his description of the work as a "bare and imperfect outline of some of the more important theorems of political economy.[3] In fact, the book does not furnish even the skeleton of a system of economics. It contains only the uneven results of Jevons's attempts to expand and correlate his brilliant suggestions of 1862. Despite his continued interest in these doctrines, the varied interests of the author and his fertile originality kept his enthusiasm constantly centered on new intellectual problems. The *Theory* as it stands was written under the pressure of a sense of unfulfilled duty to himself and to his own reputation. Jevons's intellectual inde-

[1] This matter has been carefully discussed by Mr. P. H. Wicksteed, in the *Quarterly Journal of Economics*, vol. III, pp. 297–314.

[2] Fourth edition, p. xliii.

[3] Jevons to d'Aulnis de Bourouill, *Letters and Journals*, p. 309.

pendence was too rugged and his antipathy to the school of Mill too deep to permit him to fit his own theories loosely into the general body of economic doctrine by a process of easy eclecticism. Not that he did not consciously accept and utilize certain established economic doctrines, such as the theory of land rent, but rather that all such had to be wrought over in his own thinking and thoroughly amalgamated with his own theoretical departures. The book, then, is not inconsistent, but rather, as was suggested, uneven and unsystematic. At one point, for example, we find such marks of staleness as tedious mathematical elaborations of truisms in the theory of exchange; at another point a brilliant but uncrystallized theory of interest. The first four chapters are, in general, more completely elaborated than are the remaining four. No one realized these deficiencies more clearly than Jevons himself. It was his hope to complete a large systematic treatise on political economy, which should embody his own views but which should be non-mathematical and in other ways less technical than his *Theory*. The disappointing fragments of the *Principles of Economics* do not enable one to say with any confidence that the project, if completed, would have been successful.[1]

The *Theory of Political Economy* is one of those books which are more widely quoted than read. There are indications, moreover, that the preface and the first three chapters are somewhat better known than the balance of the book. It is unfortunate for both Jevons's reputation and his influence that this should be so, for

[1] The admirably lucid *Primer of Political Economy* does not throw any light upon the question of Jevons's power to weld his doctrines into an effective and rounded system of economic theory. Intended for elementary schools, it contains little theory, and what there is is mostly of a conventional sort.

the last three chapters contain the outlines of his theory of distribution, which, fragmentary as it is, is by no means a negligible contribution to economic science. Even the better known parts of the *Theory* are worth a reconsideration at this time, if for no other reason than that new currents of economic thought have brought Jevons's work into a somewhat different perspective than that in which it appeared to his contemporaries.

Jevons was the first significant writer consciously to blend English utilitarianism with the theories of abstract economics. And no subsequent writer has more unqualifiedly and definitely accepted the hedonistic interpretation of economic motives. So extended is his discussion of the calculus of pleasure and pain and so explicit is his reliance upon it, that it might easily be inferred that the significance of his whole theory hinges upon the adequateness of the hedonistic psychology. But such, I think, is not the case. Mr. Wicksteed, himself a disciple of Jevons, has clearly shown in his *Common Sense of Political Economy*, that the notion of "marginal significance" retains as much validity when instincts and habits are counted among the forces governing men in their economic relations as when only "economic men," actuated solely by a reasoned pursuit of a maximum of pleasure, are postulated. Pareto, although many of his theorems (especially those relating to the *maximum d'ophélimité*) are, like those of Walras, essentially hedonistic, is careful to point out that his general theory of "economic equilibrium" does not rest upon the postulate of utility (*ophélimité*).[1] Jevons's theories of exchange and distribution could be similarly divorced from their apparently hedonistic basis without

[1] *Manuel d'économie politique*, p. 169, note.

substantial alteration of their essential features, and with a distinct gain in the flavor of actuality.

Jevons's elaborate exposition of the theory of utility was not, however, without purpose. The *Theory* embodies a protest against the economics of Ricardo and Mill. The chief count in the indictment was their neglect of the factor of utility in the "utility and scarcity" couplet, and it devolved upon Jevons to give this factor what he thought was its rightful position of primacy. But his interest in the analysis of utility went farther than establishing it as the psychological basis of the theory of catallactics. In Jevons's view economics and ethics were inseparably connected. "The object of economics," he says, "is to maximize happiness by purchasing pleasure, as it were, at the lowest cost of pain." [1] Jevons seems to have hoped that the principles of economics would be useful guides to individual as well as to social action. This, of course, would involve a curiously circular system, for his general economic principles were supposedly deduced only from assumptions as to the motives which do actually operate in economic intercourse. But Jevons's utilitarian ethics could be eliminated from his work even more easily than his hedonistic psychology. Whatever would have been the position of ethics in his completed system, the book as it stands is confessedly concerned only with "the mechanics of utility and self-interest." [2]

The most famous contribution of the *Theory* is the concept of the "final degree of utility." This concept is not precisely like the concept of "marginal utility,"

[1] *Theory*, p. 23. Compare the frequent assertions of the ethical purpose of economics in the *Primer*.

[2] *Theory*, p. 21.

with which it has usually been identified. For practical purposes both concepts come to about the same thing, and such difference as there is may be attributed to the fact that the concept of "marginal utility" was not, in its origin, formulated mathematically.

The final degree of utility is, substantially, the quotient of marginal utility (conceived as the utility of the marginal increment) divided by the size of the marginal increment, where this increment is very small. More accurately, it is the *ratio* of the increase in total utility to the increase in the quantity of the commodity *at the margin.* "Final degree of utility" is not only the more precise notion but it has the further advantage of being conceptually independent of the nature of the unit (pounds, bushels, yards, etc.) in which the commodity in question is usually measured. This latter quality made possible its convenient use in Jevons's "equations of exchange." [1] But marginal utility is the less abstract concept, and is undoubtedly better adapted to popular exposition.

Jevons did not utilize the notion of "subjective value" in his theory of exchange, although he suggested that final degree of utility is synonymous with value in the sense of "intensity of desire or esteem for a thing." What Jevons, in common with other mathematical economists, primarily concerned himself with was the ratio of exchange. His fundamental theorem is that the ratio of exchange of two commodities is inversely *proportional*

[1] The theorem that a person tends to adjust his expenditures so that the "marginal utilities" of the various commodities consumed are equal is, of course, true only when it is stipulated that the marginal increment is conceived as the amount of a commodity that can be bought with a dollar or other small unit of money. "Marginal utility" is, for this reason, poorly adapted to the analysis of barter. "Final degree of utility" is the precise equivalent of Pareto's *ophélimité élémentaire* and of Walras's *rareté.*

to their final degrees of utility. Final degree of utility is thus used, not as measuring value, but as determining a proportion.

The mathematical statement of this principle took the form of the well-known "equation of exchange," which remains Jevons's most substantial contribution to distinctly mathematical economics.[1] It cannot be said, however, that this equation has proved itself a useful tool in economic analysis. It assumes direct barter between a pair of traders. Professor Edgeworth has shown in his *Mathematical Psychics* that, given under these conditions only the functions which express the utility to each of the traders of the commodities to be (partly) exchanged, the ratio of exchange will be indeterminate. Jevons meets the difficulty by invoking the "law of indifference" (that there cannot be more than one price for any one article in the same market at the same time). This means, of course, that his traders are not to be considered as isolated, but as members of a general market. There are, however, substantial reasons for the thinking that postulating the "law of indifference" is equivalent to assuming the existence of a general medium of exchange.[2] Jevons did not realize that he had implicitly made such an assumption. Like many other economists he seems to have considered purchase and sale and money prices as only the superficial aspects of barter and of direct "ratios of exchange." By thus neglecting to deal directly with the positive facts of the market, he forfeited something of scientific precision as well as some advantage in exposition.

[1] It is, for example, the only formula in the *Theory* mentioned by Pareto in his article "Anwendungen der Mathematik auf Nationalökonomie," in the *Encyclopädie der mathematischen Wissenschaften*. The equation is identical with Walras's "equation of maximum satisfaction."

[2] Cf. above, p. 202.]

When we pass from the bargains struck by pairs of traders to the larger problem of the determination of the ratios of exchange in the general market, the difficulties of Jevons's methods are multiplied. Jevons avoids rather than meets these difficulties by utilizing the concept of the "trading body," which "may be a single individual in one case; it may be the whole inhabitants of a continent in another; it may be the individuals of a trade diffused through a country in a third. England and North America will be trading bodies if we are considering the corn we receive from America in exchange for iron and other goods. . . . The farmers of England are a trading body when they sell corn to the millers, and the millers both when they buy corn from the farmers and sell flour to the bakers." [1] The trading body, whatever its nature, is made to play the rôle of an individual trader. Thus, by the equation of exchange, the ratio at which English iron is exchanged for American wheat is the reciprocal of the ratio of the final degree of utility of iron to the final degree of utility of wheat, which latter ratio must be the same for both trading bodies. Jevons was too clear a thinker really to adopt the vague concept of "national (or group, or social) final degree of utility." But he thought that the economic laws representing the conduct of groups might be thought of as "fictitious averages" of the laws representing the conduct of the different individuals in the group.

The highly abstract and figurative nature of the concept of the "trading body," as applied to the whole group of dealers in a commodity, is indicated by the fact that if it were taken in any literal sense the market could not be supposed to be competitive. With all the millers and all the bakers in England conceived rigidly as a

[1] *Theory*, p. 88

single pair of traders, the "law of indifference" could not be invoked, and the equation of exchange would not lead to a determinate ratio of exchange. Jevons's refusal to assume a general medium of exchange is primarily responsible for these difficulties. When the existence of money is taken into account Jevons's equation of exchange leads very naturally to the analysis of *supply and demand at a price.* In discussing Thornton's criticism of the "laws of supply and demand," Jevons for the moment slips into the vocabulary of the money economy: "Any change in the price of an article will be determined not with regard to the large numbers who might or might not buy it at other prices, but by the few who will or will not buy it according as a change is made close to the existing price." [1] Jevons fails to see the inconsistency of this now generally received principle of the significance of marginal traders with his own notion of the operations of a trading group as an average of the operations of its members, and its perfect consistency with the theory of market value for which he takes Mill to task.

Jevons's discussion of the relation of the costs of production of commodities to their ratios of exchange is also based on the hypothesis of a barter economy, and is subject, in general, to the same limitations that obtain in his treatment of the relation of utilities to ratios of exchange. The ratio of exchange of two commodities is held to be directly proportional to the "degrees of productiveness of labor applied to their production." The "degree of productiveness of labor" is the ratio of product to the labor expended in producing it at the margin where the degree of disutility of labor is equal to the

[1] *Theory*, p. 109. In the *Primer*, p. 100, he uses the very "equation of demand and supply" for which he criticizes Mill in the *Theory*, p. 101.

degree of utility of the product. Costs of production (in labor) are, of course, inversely proportional to the "degrees of productiveness."

In this manner Jevons develops new equations of exchange in which cost of production plays precisely the same rôle assigned to utility in his previous equations. His argument that the significant fact, however, is utility rather than cost of production, opened up a whole field of controversy which need not be surveyed here.[1] It may be noted, however, that Jevons's position on this question is completely dissociated from his general theory of exchange, and the reasons he gives for his attitude are extraneous to the general run of his analysis. Furthermore, in his brief for the dominance of utility he breaks with his resolution to use the term "ratio of exchange" instead of "value" wherever possible, and his argument seems to be somewhat dependent upon the dubious meaning of the latter word. Finally, it is obvious that his criticism of the "cost-of-production theory of value" had no bearing upon the prevalent form of that theory, which was simply a statement of long-period price tendencies under conditions of free competition.

The central point in Jevons's theory of distribution is his doctrine of interest. It is an indication of the comparative neglect of the latter portions of the *Theory* that only one writer,[2] so far as I know, has credited Jevons with the marginal productivity theory of interest. Not even the similarity of the diagram[3] which Jevons uses to illustrate his theory to those used in recent exposi-

[1] So far as the validity of Jevons's own arguments is concerned, Marshall's criticism in his *Principles*, fifth ed., Appendix I, seems to me to be definitive.

[2] G. Cassel, *The Nature and Necessity of Interest*, pp. 52–55.

[3] *Theory*, p. 258.

tions of the marginal productivity theory seems to have been noticed. In fact, however, there is an important difference between Jevons's statement of the theory and its more recent formulations. Considering the principal function of capital to be to extend the interval of time "between the moment when labor is exerted and its ultimate result or purpose accomplished," [1] he defined capital as the "aggregate of those commodities which are required for sustaining laborers of any kind or class engaged in work." [2] Capital may be "invested in" factory buildings, machines, and other auxiliary instruments of production, but the "stock of capital" is the stock of sustenance. Consequently interest appears as created by a larger product, got when a given amount of labor is distributed through a longer period of time. Moreover, he assumed, without proof, that the product for the same amount of labor "varies as some continuous function of the time elapsing between the expenditure of the labor and the enjoyment of the result." It is to the period of the investment, then, rather than to the amount of the capital invested that he attributes diminishing productivity.

It should be remembered, however, that here, as in his theory of exchange, Jevons is endeavoring to dig below the surface of our money economy. His "investment of capital" is a social process, and not a matter of the expenditures of individual entrepreneurs. The common form of the theorem of diminishing productivity rests upon the analysis of entrepreneurs' costs. By a somewhat generous interpretation Jevons's theory might be said to imply substantially the facts that are utilized in the modern form of doctrine.[3] On the whole,

[1] *Theory*, p. 228. [2] *Ibid.*, p. 223.

[3] An attempt to show this agreement has been made by H. S. Jevons

however, his theory bears a closer relation to Böhm-Bawerk's doctrine of the "technical superiority of present goods" than to current marginal productivity theories.

Other features of Jevons's theory of distribution need less consideration. He accepts the orthodox theory of rent and propounds a residual claimant theory of wages. He seems to have been led to this latter doctrine by his marginal productivity analysis. Using labor as the fixed and capital as the variable factor, wages naturally appeared as a surplus or residuum on the product of all but the final increment of capital. The possibility of treating capital as the fixed factor and labor as the variable seems to have escaped his notice.

Some general aspects of Jevons's use of the mathematical method remain to be considered. The book is probably the best known single brief for the use of that method. But the work itself is mathematical only in a superficial way. Except for its use of mathematical symbols it is, for the most part, mathematical only in the sense that any economic reasoning dealing with changing quantities and ratios is *ipso facto* mathematical. Cournot (to take the best exponent of the possibilities of the method) entrusted himself to his symbols, and by means of strictly mathematical processes reached some conclusions which were neither obviously implied in his premises nor likely to be easily discerned by the ordinary

in Appendix I to the present edition of the *Theory*. A previously unpublished fragment on capital intended to form part of the unfinished *Principles of Economics* is printed as another appendix. Its most interesting innovations are the substitution of "capitalization" for "capital" and the implication that interest attaches only to free capital, fixed capital getting its reward in the form of rent or (as in the case of investments for a man's education) in wages. By "capitalization" Jevons means the amount of the capital multiplied by the period of investment.

processes of non-mathematical reasoning. There is no question but that some of Jevons's fundamental concepts presented themselves to him as mathematical quantities. But his manipulation of these concepts is for the most part non-mathematical. Jevons was not an accomplished mathematician. In some places the awkwardness of his mathematical processes [1] indicates that he is giving a mathematical garb to results reached by non-mathematical reasoning. Such attempts as he makes to develop some of the mathematical possibilities of his concepts are perfunctory. His use of the differential calculus is more apparent than real. "Final degree of utility," for example, although conceived as a derivative, is treated as an algebraic ratio. There seems to be no reason for questioning Marshall's judgment that "the book would be improved if the mathematics were omitted, but the diagrams retained." [2]

It is difficult to estimate the amount of Jevons's influence, because it is impossible to dissociate it from that of Walras, Sax, Menger, Böhm-Bawerk, Marshall, and others. In England the line of filiation runs clear only in the case of Professor Edgeworth, who has taken Jevons's work for the point of departure of some of his own brilliant developments of mathematical economics, and of Mr. Wicksteed, in whose work we find a rare degree of originality coupled with a faithful adherence to Jevons's general point of view and to some of his leading doctrines. In Italy, where a translation of the *Theory* appeared in 1875, Jevons wielded for some time a considerable influence, but in later years the influence of the Lausanne School seems to be dominant (so far as mathe-

[1] Cf. especially pp. 127–34 of the present edition.
[2] *Academy*, vol. III, p. 132.

matical economics is concerned). In Germany, Scandinavia and Holland, individual writers such as Launhardt, Auspitz and Lieben, Wicksell, Cassel, and Pierson are to be counted among those who have followed Jevons at one point or another. In the United States, of course, Jevons's influence has been quite overshadowed by that of the Austrians.

It is clear, however, that although Jevons did not bring about the revolution in economic theory which he desired, the current body of economic doctrine is measurably different from what it would have been if Jevons had not written. His doctrines have been absorbed into the general structure of economic theory and reconciled with the political economy of the Ricardian line in a manner which we may suppose Jevons would have neither imagined nor wished. But the position of the *Theory* as one of the four or five great books of nineteenth-century English political economy is secure. It retains in a surprising degree the quality of making a fresh and vivid appeal to the reader's interest. Its transparent intellectual honesty and the entire absence of scientific pose also commend it. If only by the dissent which it may compel, it still directs the attention of the thoughtful reader to the fundamental problems of economic theory.

XII

THE TREND OF ECONOMICS, AS SEEN BY SOME AMERICAN ECONOMISTS [1]

"It will be said, I suppose," Professor Tugwell observes in his Introduction, "that this book is a sort of manifesto of the younger generation; and in a sense it is that, though none of us, I think, has meant to put first in his writing the critical attitude."

I suspect, indeed, that the first impression most readers will get from the book is that its authors take a distinctly critical, or even condescending, attitude toward the achievements of the earlier generations of economists. Professor Mitchell finds their discussions "dull" (if taken literally), and their conclusions "dubious." Professor Slichter holds that "Orthodox doctrine is formulated with as little reference to machinery, science applied to industry, corporations, trade unions, the credit system, commercial and investment bankers, as were the theories of Adam Smith." Mr. Soule alludes to "the barrenness of premature senescence which has existed in economic science." Professor Tugwell is quite as severe: "Economics incontestably has got a bad metaphysical odor that only a renaissance of rebuilding from the ground up can dissipate."

But a more careful examination reveals that these extreme views are exceptional. Outside of the chapters

[1] A review of *The Trend of Economics*, by M. A. Copeland, S. H. Slichter, F. H. Knight, A. B. Wolfe, W. C. Mitchell, P. H. Douglas, R. G. Tugwell, F. C. Mills, W. E. Weld, R. T. Bye, J. M. Clark, R. L. Hale, and George Soule. Edited, with an Introduction, by R. G. Tugwell. New York, Alfred A. Knopf, 1924. [Reprinted from the *Quarterly Journal of Economics*, vol. XXXIX, No. 2 (February, 1925).]

contributed by the writers who have just been quoted, there appears to be little disposition to raise the red flag of scientific revolution. All shades of opinion are represented. There are radicals, there are conservatives, and there are some who occupy a middle position. I shall find it convenient, in surveying these varied chapters, to follow a path that runs generally, though not consistently, from the right toward the left.

Professor Knight's chapter on "The Limitations of Scientific Method in Economics" would give distinction to any book. Because I put Professor Knight among the conservatives, it should not be inferred that he undertakes a general defense of prevailing economic doctrines. He is conservative merely in that he insists that the credentials of all projects for the radical reconstruction of economics be submitted to careful scrutiny. He is not willing to follow a new way just because it happens to be the mode. But his attitude appears to be without prejudice. The old methods and the new have to measure up to the same logical standards. In short, Professor Knight thinks for himself. Moreover, his knowledge of psychology and of the methods of science equips him adequately for his task.

The general run of his argument is somewhat as follows. Science is instrumental, concerned with control, with means rather than ends. It supplies, in effect, an elaborate technique of prediction. Yet it is of small help in many of the practical decisions of life. It fails most notably in the prediction of human behavior and of social changes.

Behaviorism fails, first, because in rejecting consciousness it rejects a *useful* datum. Science itself finds it convenient to retain an element of animism, and

science, as a part of intellectual life, is built on intercommunication, which implies the recognition of the consciousnesses of others. Behaviorism fails, second, because the reactions of different persons are not alike. For one thing, no two of them have passed through precisely the same experiences. There is no stable relation, common to all persons, between stimulus and response, and the stimuli themselves are often responses to other stimuli. "The childish doctrine of instincts" is also valueless for prediction. Innate propensities are general, not specific.

In general, in dealing with human behavior, whether consciousness is or is not taken as a datum, it is hard to lay hold of "objectively recognizable features in the situation which are uniformly associated with different possible eventualities." So far as the prediction and control of human behavior is at all possible, it is not by discovering and applying rules. The problem belongs to the field of art, of interpretation and suggestion, not to the field of science.[1]

[1] I suspect that Professor Knight carries his skepticism on this point too far. Certain "principles" of advertising and of "salesmanship" seem to have been developed. But doubtless the largest achievements in either field will always demand a more sensitive and pliant knowledge of human nature than can be formulated in rules. Sir Frederick Pollock said of Machiavelli's influence, "no man ever learned the secret of despotism out of a book." It should be remembered, however, that men's behavior (or, better, their conduct) can be modified by increasing their knowledge as well as by playing upon their weaknesses. Present-day proposals for "controlling human behavior" so as to moderate the cyclical fluctuations of business, rest on the assumption that men will make *rational* use of increased knowledge respecting the outcome of different types of business policies. In a similar way J. S. Mill and other Malthusians proposed to improve the economic status of wage-earners by acquainting them with the relation between the increase of the population and the movement of wages. Neo-Malthusians likewise put their trust in education, but assume, with reason, that knowledge that bears directly upon present personal interests is more likely to be used than knowledge that affects merely the remote interests of a class.

So also with the study of history and of institutions. History gives judgment and insight, it does not give rules. "The useful knowledge of history is chiefly unconscious knowledge." There are certain fields, Professor Knight admits, where *statistical* laws are possible. But it is difficult to get the facts; the method is applicable only to the "elemental" factors of life, and the behavior even of groups is unstable and unpredictable. Groups as well as individuals "grow in historical isolation." [1]

Is there, then, any room left for a *science* of economics? Professor Knight holds that there is. "In a limited field in economic data, due largely to the fact that exchange has reduced the factors to definitely measurable quantities, we can have an exact science of the general *form* of relations." In a large sense "economizing" covers about all of rational behavior. But through a division of labor among different sciences, "economic theory has in practice come to be restricted to the analysis of social interaction and coördination through the price mechanism. . . . It can tell us little in the concrete, and its chief function is negative — to offset as far as possible the stupid theorizing of the man in the street. The real sociology and economics must be branches of literature as well as of science."

Professor Wolfe contributes a chapter on "Functional Economics." Under this forbidding title he

[1] Such, it will be remembered, was the ground of Knies's denial of the possibility of "scientific" laws of economic life. I think that Professor Knight does less than justice to the possibilities of the statistical method. Practical difficulties are being surmounted, the range of statistical information is being increased, technique is being perfected. Progress along these lines in the future promises to be more rapid than in the past. Furthermore, statistical groups are generally "isolated" only in part.

writes about the place of science in economics and about a number of other things. Professor Wolfe is particularly opposed to the "pure science ideal," that is, that "economics is to be the virgin pursuit of knowledge for its own sake." There are, he holds, two other tenable ideals. First, research, while remaining objective, may be directed to problems that have a practical or ethical interest. Second, research may be "undertaken from ethical motives as well as from intellectual curiosity, and with some fundamental ethical norm or standard in mind." This second alternative is the one Professor Wolfe prefers. Economic research so motivated gives us "functional economics." It is not quite clear why he should worry about economists' motives, for he professes to be an out-and-out determinist.[1]

Be that as it may, it is clear that the winds of pragmatism and instrumentalism have passed him by. "Pure science" is a logical fiction, although it is reputed to have been pursued at the Academy of Laputa. Knowledge not related to some human interest is not knowledge. Never does an investigator seek merely what he may find, but always an answer to a question. Applied science differs from what is called pure science only in its lesser degree of generality, in the immediacy and particularity of its problems. It is a commonplace that in the history of the physical sciences the most fruitful and, in a sense, the most "practical" researches have been those which have been free from the hampering specifications which immediate and particular problems

[1] "A scientific psychology must be deterministic. . . . Man's behavior is as mechanical, in the last analysis, as that of any other part of nature." On the other hand, "someone is going to attempt the task of constructing a factually supported theory as a guide to rationally directed social or economic evolution, toward an organization more effectively serviceable to worthy human ends than the one we now have." Surely this is riding two horses at once.

impose. In economics as in economic life it is likewise true that roundabout methods have often proved to be the most productive.

The proper placing of ethics in its relation to economic inquiries is a matter of technique, not of principle. An economic inquiry may be colored and shaped by an ethical interest, even when the inquiry is kept free of formal ethical postulates.[1] Economics may be made to deal only with "what is," as distinct from "what ought to be"; and yet the particular report it makes, the precise aspects of "what is" to which it attends, may be determined by an interest in "what ought to be." But economics would lose rather than gain by an alliance with formal or "scientific" ethics. Professor Wolfe's "functional welfare economics" is to have as one of its fundamentals "a generally accepted, psychologically grounded, norm of welfare." This requirement is a heavy handicap. In fact, it puts "functional economics" pretty definitely out of the running.

Professor Bye, who writes of "Some Recent Developments of Economic Theory," defends the position which Professor Wolfe attacks. He repeats the old arguments for "pure science," for a "purely descriptive economics," and for its sharp separation from "applied economics." "Knowledge for knowledge's sake, without any intruding ideas about welfare, is the surest way toward the ultimate promotion of economic well-being." This position, I should say, involves a confusion. Science may perhaps be unbiased. But science can never be concerned with "pure description," for the reason that description is never "pure." Scientific as well as æs-

[1] This, I suppose, is substantially what Professor Wolfe has in mind as the second of the three possible ideals of economic science.

thetic description requires selection and valuation. Even bias, I suspect, is not without its uses, if only as an energizer, a catalytic agent.

The general significance and the present status of the theory of value and distribution are the topics around which Professor Bye's discussion centers. If he says little that is new, he says much that is judicious. And if he fails to dig very deep, it is because he is excavating in the shallow débris of the dialectical controversies with which, during the past thirty years, too many American theorists have occupied themselves. I get the impression that if he were to deal with substantive problems, involving more than merely verbal issues, Professor Bye could do work of a high order.

Professor Clark has never done anything better than the first two sections of his paper on "The Socializing of Theoretical Economics." In them there is a rare combination of felicitous expression and analytical finesse. Professor Clark's conclusions are in some respects different from those of Professor Knight. But their contributions to this volume have one thing in common. They show that both authors have given serious attention to modern developments of logical theory, and have woven some of those developments into their own thinking.

Professor Clark's general view of economics is pragmatic. But he does not hold that the problems of economics are merely problems of "economic efficiency." Ethics might conceivably be ruled out, the notion of efficiency be kept distinct from any avowed notion of what is good, but only at the cost of introducing bias by including or emphasizing certain values rather than others.

Selection appears not only in the choosing of pro-

blems, but in picking the ways and means of attacking them. The materials and methods that satisfy the canons of scientific adequacy impose their own limitations upon us. "Yet," Professor Clark judiciously observes, "it is probably impossible to invent a system of logic so tautologically sterile, or a behaviorism so aridly descriptive, as wholly to prevent the human mind from getting at the inner significance of facts." But if economic science is to make real progress, he contends, deduction and induction must be consciously used, so that each shall give significance and meaning to the other.

Economic theorems that are universally true are mere truisms, tautologies. For example, "We are told that price is fixed at the point where supply and demand become equal." But, "It is only as common sense or induction adds descriptions of the behavior of 'supply' and 'demand' that the proposition has meaning attached to it, by a sort of associative process, and becomes a vehicle or symbol of truths which its literal formulation does not categorically denote."

In determining selection and emphasis, Professor Clark suggests, time and place are of primary importance. A proposition about the existing economic order necessarily hinges upon the character of the substitutes for that order which are "within the bounds of serious possibility." Further, "it is probably true that economic conceptions have mostly been framed with reference to particular errors or evils which they might seem to combat." Smith's economics was *directed* against mercantilism. The theory of marginal productivity was an answer both to Marx and to "wage-fund extremists."

Thus far I can follow Professor Clark with confidence and with grateful approval. But I cannot follow him in his next step, in which those partial aspects of truth

which are marshalled for practical or even for militant purposes are made to play the rôles of opposites, in a kind of Hegelian dialectic. Competing economic theorems are generally complementary, sometimes overlapping, but never (unless there is positive error on one side or the other) opposites.

Professor Clark's experiments in non-Euclidean economics are undeniably clever, but they do not convince me. He examines eight "eminently respectable propositions," and manipulates them by processes which he lumps together under the name of "inversion" but which are really of several different types. Some of his theorems, in fact, are not "economic" propositions.

Two writers, Professor Mills and Professor Weld, deal with the possibilities of inductive methods. Professor Mills's paper "On Measurement in Economics" is a workmanlike production. Nevertheless, I believe, some of his principal conclusions are untenable. That he should have made some slips is not surprising, for he traverses ground where the footing is treacherous. The present status of the logic of statistics is unsatisfactory, and there is room for differences of opinion on fundamental matters.

Professor Mills starts with certain considerations suggested by Merz's chapter [1] on "The Statistical View of Nature," and by Josiah Royce's notable paper on "The Mechanical, the Historical, and the Statistical." [2] He

[1] J. T. Merz, *A History of European Thought in the Nineteenth Century*, vol. II, chap. XII.

[2] *Science*, new series, vol. XXXIX, p. 551. "The object of statistical knowledge is not the single [historical] event and is not the invariant [mechanical] law, but is the relatively uniform behavior of some average constitution, belonging to an aggregate of things and events, and the probability that this average behavior will remain, within limits, approximately, although always imperfectly uniform." The statistical

is led to generalize respecting the "statistical character of useful knowledge." The difficulty with "the mechanical conception," Professor Mills holds, is that it "involves the notion of sameness, of the precise identity of all the things or events described by a given law." This is a dubious generalization.[1] But taking it as it is, what of it? Neither the physical scientist, *qua* scientist, nor the economist, *qua* economist, is concerned with the ultimate constitution of knowledge. The practical problem is one of getting useful and reliable ways of formulating and organizing knowledge.[2]

"A law as the scientist sees it," observes Professor Mills, "does not govern." No, but the notion that it does govern served a very useful purpose in its day.

view, Royce held, is not a mere makeshift substitute for the mechanical view. He argued that since verification is always inexact, an approximate theory is better than an exact one. "There is good reason to say that the statistical force is the canonical form of scientific theory." Of course this is nothing more than interesting conjecture. The difficulty is that statistical knowledge, for aught we know to the contrary, might be resolved into the mechanical form, while important types of mechanical knowledge appear to resist statistical formulation. This particular range of problems has never been more ably discussed than by Cournot, especially in his *Essai sur les Fondements de nos Connaissances*, although his conclusion that the world itself, like our knowledge of it, is both mechanical and statistical, has never gained a following. Cournot's distinction between historical and scientific knowledge was essentially like the one drawn by a number of modern logicians, notably H. Rickert. Royce's view was less fully developed. Professor Mills, it may be observed, follows Maxwell rather than Royce when he asserts that "mechanical laws assume the historical type of knowledge." He appears to mean that mechanical knowledge assumes *complete* induction — surely not a tenable position.

[1] What the mechanical view really assumes is that the *relations* of *quantitative* aspects of things are stable, in the sense that their variations may be described by mathematical functions. There are other necessary assumptions, but only such as are equally required by the statistical view.

[2] Professor Mills urges that the data of economics are more uncertain than those of the physical sciences. But this puts additional difficulties in the way of the empirical as well as of the rational formulation of tendencies.

Professor Mills follows Marshall (and J. S. Mill) in holding that an economic law is a statement of a "tendency." [1] Now the notion of a "tendency" in phenomena is quite as animistic as the notion that "law governs." But I should not reject it on that account. It is a convenient way of indicating with a minimum of circumlocution that a law is a general statement that has prediction value.

I cannot, however, follow Professor Mills in holding, with Royce, that "tendency is an essentially statistical term," "a statement of average relations." That a projectile tends to follow a path defined by a curve fitted by least squares to the observed paths of different sorts of projectiles, selected at random and discharged under haphazard conditions, is not a useful and not even an intelligible statement. That a projectile tends to follow a path determined by accurate observations of the behavior of similar projectiles, discharged under generally similar conditions is, on the other hand, a useful statement which serves as a practical rule in gunnery. That a projectile tends to follow a path defined by a parabola is also a practically useful statement.

That the prices of agricultural products tend to increase as population grows is, under present conditions, a rational (mechanical) rather than an empirical statement. It is in no proper sense an "average." That wages tend to lag behind advancing prices is, let us say, an inference based in large part upon an induction, but it is vastly more than an "average." A statement of

[1] Professor Mills prints two graphs (on pages 52 and 53), one illustrating physical and the other economic "law." Both show relationships of physical variables. In a rigid sense both are illustrations of diminishing returns. Their only differences are that the second "law" has *more* economic significance than the other, and that in its case the observed facts are not summarized so well by a smooth curve.

historical fact is not a statement of tendency. An average, as well as a single event, may be unique.[1]

In the field of mechanics, which Professor Mills cites by way of analogy, the real cleavage is not between rational mechanics and statistical mechanics, but between rational mechanics and empirical mechanics.[2] But

[1] The larger the purely "statistical" element, the more, that is, an average enters into a process purely *as* a statistical average, rather than as a device for adjusting observations on a fixed but unknown magnitude, the greater is the uncertainty of prediction. This distinction is not merely verbal. To say that a statistical inquiry can be brought under the general theory of errors of observation is to say, in effect, that it supplements (and is supplemented by) a large body of relevant experience, organized in rational or mechanical form. The same principles are involved in our habit of approving statistical findings that appear to be "reasonable." In practice it is generally some such joining of statistical and rational considerations that makes it possible to get workable results out of what appear on the surface to be weak or meaningless a posteriori probabilities. Urn schemata are inadequate representations of the universe of experience within which the statistician works.

[2] The kinetic theory of gases, as developed by Maxwell, Boltzman, and Gibbs, is the canonical example of "statistical" mechanics. The method is a priori, not empirical, and serves as a means of *explaining* experience, not of summarizing or describing it. Conceivably, making reasonable suppositions with respect to the position and movement of each molecule, the whole mechanism could be explained by a system of equations. Millions of equations would be necessary, so that approach is cut off. But the number of molecules is so large, that it *makes no practical difference* whether they be considered as bound by law or free. In either case the fundamental assumptions are Newtonian. Gibbs himself never suggested that his formulas gave a clue to the nature of knowledge. He was content that they gave results that "coincide sensibly with the general laws of thermodynamics." Similarly the theorem that entropy (useless energy — playing a part in mechanics curiously like that of utility in some systems of economics) tends toward a maximum, may be formulated as a statement of a tendency toward change from a less probable to a more probable state. Here again statistical mechanics affords merely an alternative *form* of explanation. It gives no hint respecting the constitution of the data of experience. (For a good elementary account of these matters see Emile Borel, *Le Hasard*, chaps. VI and X.) On the opposite margin of the science of mechanics, remote from these bold generalizations, we find empirical formulas, "statistical" in the sense that they involve averages, which have not been reduced to rational rules. Here again economy is the controlling consideration. The "uncertainty" of statistical mechanics is not the uncertainty of empirical formulas. Royce confused two different things.

rational mechanics is merely a way of organizing experience, while empirical mechanics is fairly sure to be rationalized whenever that process is economical. Experience and theory go hand in hand, or rather, by their aid science climbs upward hand *over* hand. In economics, likewise, inductive findings will give accuracy and concrete meaning to the theorems of rational [1] economics, while theory will continue to analyze, interpret, relate, and support the results of inductive studies and of such other new experience as comes to its hand. (Professor Mills, it is proper to say, does not hold to the contrary.) The growing use of quantitative methods is the most promising development in contemporary economics. But it will prove relatively sterile if it does not lead to a renaissance of theory.

Professor Weld's chapter on "Regional Comparison and Economic Progress" does not call for extended comment. The younger economists, unlike their teachers, Professor Weld tells us, are interested in welfare. To this end "we must learn of causes or else shut up shop." Knowledge of causes may be derived from regional comparisons, so as to "make it more possible for backward nations to be brought into better alignment with modern progress." Professor Weld suggests how it might be done. The obvious comment is that one concrete achievement in this field of research, no matter how small, is worth any amount of schematizing.

Professor Weld's scheme, after all, is merely that the researches he proposes should be guided by Mill's canons of experimental inquiry. It is surprising that Professor Weld should have failed to note Mill's emphasis upon the difference between sciences of mere observation and

[1] I trust that it is still possible to use this adjective with a logical rather than a psychological implication.

sciences of experimentation. It is even more surprising that he should not have noted that Mill devoted a chapter largely to the task of showing that his canons of experimental research could not safely be used in regional economic comparisons.[1]

The title of Professor Tugwell's paper, "Experimental Economics," suggests that he too is writing of inductive methods. It suggests also that he too confounds induction and experimentation. Happily, the paper belies its title. One's judgment of its quality must depend largely upon the type (as distinct from the severity) of the standards by which one rates it.

It is plentifully besprinkled with the various tags of modernism. Dogmatic opinions are thrown around recklessly. There is no evidence of any effort to achieve logical consistency.[2] Yet what Professor Tugwell has to say is not to be dismissed lightly. Impressionistic as it is, his paper has a distinctly interesting intellectual back-

[1] Mill undoubtedly went too far in his skepticism. In the eighth edition of his *Logic* he appended to his general discussion of plural causes (Book III, chap. X) a note which begins as follows: "It is justly remarked by Professor Bain, that though the Methods of Agreement and Difference are not applicable to these cases, they are not wholly inaccessible to the Method of Concomitant Variations." But he made no change in his chapter on the use of inductive methods in the social sciences (Book VI, chap. VII). Not only did Mill write when statistical technique was relatively undeveloped, but it is clear that he had particularly in mind a special problem (the effects different commercial policies have on national prosperity) where the nations that might be compared are few and the effects to be detected ill-defined and relatively slight.

[2] I select only one example. Professor Tugwell holds that economics has suffered because its development has largely been at the hands of teachers whose task is "to communicate to students the accumulations of scientific knowledge that society has been able to build up." Unfortunately for this generalization certain important exceptions would occur to any one. So, according to Professor Tugwell, Ricardo, Marx, and the Mills, although "they had no formal University connections, were nevertheless pedagogues." From which I infer only that Professor Tugwell does not like their doctrines.

ground. It has also a certain eager and buoyant quality. And it is better to throw logic to the winds than mess around with it blunderingly. To adapt a phrase from Professor Knight, the field of economic criticism has room for literature as well as for logic.

A paper made up of appreciations (and depreciations), with æsthetic rather than logical unity, cannot be summarized in a few paragraphs. But the picture Professor Tugwell paints is, in outline, one of a world confronted by momentous new problems, handicapped by an economics that is rigid and inflexible because the notion of natural law permeates it through and through. "Economic laws (as now formulated) are not natural laws just because they are not useful in the sense that natural laws are useful."[1] We need a more flexible, experimental, approach to economic problems. Not only what cannot be but what might be should engage our active interests. We need, in effect, a larger use of the planning and creative imagination. But along with that we need to verify and reject. Social experimentation is costly. "If we can somehow simulate experience as the chemist simulates in his laboratory, we shall have made a big advance in social methods."[2]

I should accept this in part, not as an account of what we ought to do, but as a helpful account of the actual processes of constructive thought. But I do not like

[1] Professor Tugwell cites the law of diminishing returns. Does he mean that the "experimental" economist would disregard that law?

[2] The chemist does not "simulate" experience; he creates it. And whether in our own thinking we *can* "somehow simulate experience" is an idle question; for somehow we do. Professor Tugwell quotes John Dewey on "the place of active and planning thought within the very processes of experience," but seems not to have caught the full bearing of the words. For an illuminating statement of what I take to be the core of Professor Tugwell's thesis see William James, *Some Problems of Philosophy*, p. 222.

Professor Tugwell's implications that experimental thinking affords in some way a new opportunity for economists, that past generations of economists had minds that were differently constructed, and that the notion of economic law has always been the peculiar property of the standpatter, never of the man who looks forward.[1] While he grants freedom of experimental thinking to himself and to his contemporaries, he assumes that the older economists were bound fast by tradition, by a superstitious faith in economic law, and even by the play of economic determinism.[2] I do not see how one who looks backward through smoked glasses can look forward with open and clear eyes.

Professor Slichter is quite as dissatisfied with the present status of economic theory as Professor Tugwell is, and the problems with which he thinks theory should occupy itself are like some that Professor Tugwell recommends. But otherwise their papers are radically different. Professor Slichter brings to his discussion of "The Organization and Control of Economic Activity"

[1] I shall not name American economists. But I wish that Professor Tugwell and those who hold with him would read and reflect upon Mill's autobiography and J. M. Keynes's memoir on Alfred Marshall. Presumably Professor Tugwell, like Professor Mitchell, finds Mill and Marshall dull. But surely it requires only a modicum of sympathetic intelligence to allow for the inevitable shifting of interests that comes as new problems emerge and as some of the old ones, perhaps, approach a final solution.

[2] Thus, "Professor J. M. Clark's whole approach to the field of economic theory has been ethical." But, "Adam Smith's connection with the commercial life of Glasgow bore directly upon his conception of national welfare." Professor Fisher and Professor Mitchell are interested in cures for the evils of the business cycle and of unstable prices, and their theories, therefore, are "ethically oriented." Ricardo had similar interests. But presumably these, like certain of his theories which Professor Tugwell mentions, were "directly influenced by the fact that he was a city man and a stock-broker."

an extraordinarily concrete and particularizing turn of mind. He is impatient of broad generalizations. His interest is in pointing to the looseness with which such generalizations fit the facts of experience.[1] But this implies, of course, another generalization, negative in form. Thus his principal thesis is that competition is *not* effective as a regulator of economic activity. More specifically, "an interest on the part of individual buyers and sellers in discriminating against those who disregard the common welfare is lacking, not in a few scattered instances, but in many." "Furthermore, even when individuals do have an interest in discriminating, they frequently lack the information, the time, the willingness, the intelligence, or the bargaining power to do so."

I know of no more effective brief for the prosecution than the one Professor Slichter submits. At only a few points does he put new counts into the indictment; but his methods give fresh meaning to the old ones. He does not argue his case, he merely marshals his evidence in such a way that it has a cumulative effect. I venture to say that no one can read his paper without having his own thinking affected. If that be true, Professor

[1] Professor Slichter's own generalizations are not always successful. His thesis that *aggregate* wages may be decreased by increasing the number of laborers calls for some highly improbable assumptions. His contention that the scrapping of obsolete equipment has the net effect of "enforcing" a higher rate of saving is, to say the least, unproved. His further suggestion that the paying off of debts increases social capital neglects all but the most superficial aspects of the processes of capital accumulation. The fact that few industries ordinarily produce at their peak capacity does not lead, as he supposes, to the conclusion that "chronic underproduction is an inherent characteristic of free enterprise." Such a conclusion would be a just inference from a state of chronic unemployment, but not from a state of chronic overequipment. Slips like these suggest that the "extremely general and highly simplified" type of theory to which Professor Slichter objects may not altogether have outlived its usefulness.

Slichter has made a substantial contribution to economics.[1]

I have tried to get the precise drift of Professor Copeland's discussion of "Communities of Economic Interest and the Price System," but must confess failure. However, Professor Copeland is fond of such words as pecuniary, technological, profitivity, anthropomorphic, and the like, so that his interests and the influences which have helped to shape them are fairly clear. Yet Professor Copeland is more temperate than some of his collaborators in his treatment of economists in general,

[1] Professor Slichter intends that his paper should be taken primarily as a criticism of the theory of value and of the central position that theory occupies in economic thought. "Quite clearly, the question of how value works is quite as much a part of the value problem as is the question of how it is determined." I should go quite as far, and even a little farther, for I do not believe that the two questions can be separated. This is a convenient place at which to observe that not only Professor Slichter, but a number of his fellow contributors as well, appear to identify not only "economic theory," but "economics" with certain particularly thin and abstract formulations of the general theory of value and distribution. I infer even that when the "economic systems" of earlier writers are mentioned, only their theories of value and distribution are meant. Economics is condemned for its neglect of everything but general theories of value and distribution, when the critics are closing their eyes to everything else. For example, Professor Slichter forgets that not only in Alfred Marshall's work, but in the work of many American economists, there is much that can be described by what Marshall said in the preface to his *Industry and Trade:* "The present volume is occupied in the main with the influences which still make for sectional and class selfishness; with the limited tendencies of self-interest to direct each individual's action on those lines in which it will be most beneficial to others." This sundering of a certain part of the structure of economics from its *relations* is probably in part a consequence of the unnatural separation of "pure theory" and "problems" that has prevailed in the curricula of American universities. A striking illustration of what I have in mind is at hand. A reader of this volume, with no previous knowledge of the subject, could not infer either from the individual papers or the bibliography that economists are ever concerned with problems of international trade. And yet it is impossible to understand the general structure of economic doctrine without understanding how large a part problems of international trade played in shaping it.

and he has evidently endeavored to work his way through to honest conclusions. I get the impression that these conclusions, if they could be divested of their difficult terminology, would invite thoughtful consideration.

There remain [1] the papers of Professor Mitchell, on "The Prospects of Economics," and of Professor Douglas, on "The Reality of Non-Commercial Incentives in Economic Life." Professor Mitchell's contribution is accorded the place of honor in the makeup of the volume — a circumstance which helps to attract attention to the way in which some of the papers which follow (in general, the more dissentient ones) unmistakably reflect Professor Mitchell's influence. At some points, however, Professor Mitchell himself seems to have departed from his earlier views.

No economist of his generation has made more important substantive additions to economic knowledge than Professor Mitchell. He has shown how fruitful quantitative methods may be, when guided and supplemented by skillful analysis. He has given new meaning and significance to various short-time fluctuations and maladjustments in our exchange economy. This substantive work of his fits into and amplifies the general structure of economic knowledge that has been built up

[1] To keep this review within bounds, I have to forego any extended consideration of the chapters contributed by Dr. Hale and Mr. Soule. Dr. Hale, writing on "Economic Theory and the Statesman," reviews the present status of the problem of regulating railroads and public utilities, and inquires into the adequacy of those portions of economic theory which have a bearing upon that problem. I agree with his conclusion that economic theory, in its present form, does not afford the light the problem needs. Such specific proposals as Dr. Hale makes, however, are for changes in the basis of public policy, not in economic theory. Mr. Soule's views, presented in a short paper on "Economics, Science, and Art," are in some respects like those of Professor Tugwell, and at other points are like those of Professor Mitchell.

slowly and falteringly during the last century and a half. The assumptions, the modes of thought, are such as are familiar to economists. What is new is the body of concrete experience which Professor Mitchell has patiently and deftly organized and formulated in terms of general tendencies.

It is strange, therefore, that in various other writings Professor Mitchell should have done his best to destroy confidence in the soundness of the general structure of which the best of his own work seems to be an integral part. His attacks have been directed, not at the superstructure, but at its supposed foundations. His *locus standi* was supplied by Thorstein Veblen's thesis that the character of economic science has been determined by its preconceptions. "The main canons of truth on which the science proceeded were: (*a*) a hedonistic-associational psychology, and (*b*) an uncritical conviction that there is a meliorative trend in the course of events, apart from the conscious ends of the individual numbers of the community."[1] Veblen made no attempt

[1] Thorstein Veblen, "The Preconceptions of Economic Science," *Quarterly Journal of Economics*, vol. XIV, p. 242. Reprinted in *The Place of Science in Modern Civilization and Other Essays*, p. 150. I am dealing in the text, above, with Professor Mitchell's development of the first "canon." Veblen's statement respecting the second has had an equally pervasive influence. In fact, several of the contributors to *The Trend of Economics* appear to take it as a text. So far as English economics after Adam Smith is concerned, it rests only upon Veblen's adroit fumbling of the distinction between normative (meliorative) trends and specific tendencies. Veblen was able to maintain a semblance of logical consistency only by consigning practically all of physical science, except descriptive genetics (not, strictly speaking, a science) to the limbo of outmoded superstitions. On any ground the thesis is untenable. Even Smith's "system of natural liberty" could be achieved only by conscious effort. By general consent Ricardo's view of the general trend of affairs was pessimistic. Mill's famous thesis that "the distribution of wealth is a matter of human institution only" might conceivably be ruled out, as something apart from the main body of his doctrines. Consider, therefore, the concluding sentence in his discussion of the stationary state. "Only when, in addition to just institutions, the increase of

to prove or even to illustrate his thesis respecting the first "canon." It was and has remained merely a bland assertion.

That economic science proceeded on the basis of a hedonistic-associational (or any other) psychology can be proved or disproved in only two ways, one analytical, the other empirical. I shall refer to the bearing analysis has upon the matter when I come to discuss Professor Mitchell's present position. A fair empirical test might be arranged as follows. Take Hume's economic essays, Ricardo's *Principles* or James Mill's *Elements of Political Economy*, and some important modern work, say the analytical portion of Professor Mitchell's *Business Cycles*. Revise them so that neither archaisms and turns of phrasing or references to specific problems or events shall date them. Refer them to a competent psychologist, unversed in the history of economic thought or in general economic history. The psychologist is to determine solely on the basis of the psychological preconceptions the different documents reveal: (1) on what system or systems of psychology each writer bases his reasoning, (2) which wrote before and which after Bentham, (3) which wrote before and which after the "decay" of associationism and utilitarianism. In fact, a laboratory test is hardly necessary. Any student may satisfy himself by an exercise in "experimental thinking." My own opinion is that the psychologist would be able to put his finger on *no* general psychological preconceptions.[1]

mankind shall be under the deliberate guidance of judicious foresight, can the conquests made from the powers of nature by the intellect and energy of scientific discoverers become the common property of the species, and the means of improving and elevating the universal lot." It is a pity that it should be necessary to call attention to such obvious and supposedly familiar things.

[1] I suspect, however, that he might be able to put all of the documents

In 1910 Professor Mitchell gave hostages to fortune in the form of an extended and favorable review of Professor William McDougall's *Social Psychology*.[1] McDougall's allegation that political economy rested on hedonistic postulates, and his suggestion that a classification of general instinctive propensities would afford a better basis, gave support, of course, to Veblen's thesis. Professor Mitchell restated and amplified that thesis. "Economic theory," he said, "rests and always has rested, upon the concept of human nature posited by the theorist." "It is clearly unwise for them to continue trusting and using the traditional hedonistic psychology." One of the major problems before economists, he alleged, was to trace "the processes by which habits and institutions have grown out of instincts."

But by 1914 it became clear that economists might have wasted their time if they had set seriously at work to reconstruct their science on foundations more or less like those proffered by McDougall. Instincts had multiplied under Thorndike's scrutiny, Graham Wallas had put his trust in "complex dispositions," Walter Lippman had clothed some social theory in Freudian garments, Veblen's "instinct of workmanship" had turned out not to be an instinct, and a new thing, behaviorism, had made its appearance. Most of all, a number of economists had argued that economics does not require psychological postulates.[2] Nevertheless,

somewhere in the general stream of the British empirical tradition. Veblen's own work, with its picture of systems of thought, built up by the unfolding and the synthesis of preconceptions, would be more likely to be set down as Hegelianism in a modern *patois*.

[1] "The Rationality of Economic Activity," *Journal of Political Economy*, vol. XVIII, pp. 97, 197 (February, March, 1910).

[2] A list is given by Professor Mitchell in the paper cited below. I venture to add a reference to my remarks about Jevons's hedonism. (*Supra*, p. 221.)

Professor Mitchell stuck by some of his guns, and brought up new ones.[1] "It is possible that the effort to keep the study of human nature out of economic theory may break down. . . . It may be that economists will find themselves not only borrowing from but contributing to psychology. For if that science is ever to give a competent account of human behavior it seems necessary that economists should do a part of the work." But he still insisted that hedonism had "exercised a potent influence upon economics."

In the paper now before us, Professor Mitchell again appears to alter his position,[2] and, in my opinion, for the better. I quote a significant passage:

> Economics is necessarily one of the sciences of human behavior. Whether its votaries recognize the fact or not, it endeavors to show how men deal with each other in getting their livings. Now no man can possibly give an account of economic behavior without having some working notions of human nature in the back of his head, if not on his tongue. No one can lay down any proposition about business transactions without implying[3] that men have certain standard ways of feeling, thinking, and acting in their market dealings with each other. It is therefore naïve to talk of divorcing psychology from economics. . . . It is equally naïve to talk as if the economist borrowed or could borrow all his psychological notions from the psychologists. For if economics deals with economic behavior, then it must go beyond the contributions which professed psychologists make to the general

[1] "Human Behavior and Economics: A Survey of Recent Literature," *Quarterly Journal of Economics*, vol. XXXIX, p. 1 (November, 1914).

[2] Nevertheless Professor Mitchell deems it proper to say: "Our orthodox economists have a most inadequate conception of psychology — and economics also for that matter."

[3] To infer psychological generalizations *from* generalizations about business transactions is not the same as deducing generalizations about business transactions from generalizations about "the original nature of man."

science of behavior. The economist will learn more in his own field of course if he begins tilling it with sound psychological conceptions to help him; but the soundest general conceptions of "the original nature of man" and its modifications in the course of experience will be merely a starting point for his own researches into human behavior.

The psychologists and their contributions to economics are receding into the background. It might even be inferred that the (behavioristic) psychologists have more to learn from economics than the economists have to learn from psychology. The truth is, of course, that economics not only should be or is, but always has been a study of human behavior,[1] getting its data from experience, not from psychological postulates.

Professor Mitchell's explanation of the general structure of Ricardian economics, now runs, like Professor Edwin Cannan's, in terms of the special problems of the post-Napoleonic period. But this new instrumental explanation is supplemented by a renewed insistence on the importance of a priori psychological elements in Ricardian economics. These elements, however, are no longer lumped together as "hedonism" or "associationism." We are told instead that Ricardo's conclusions "rest on the broad assumptions concerning economic behavior which a thoughtful financier had found dependable." Among these assumptions are the restless desire of capitalists for profits, modified by an aversion to certain trades, the habits and Malthusian instincts of the laborers, and the Olympian passivity of landlords.

[1] Compare the judicious findings of Professor Z. C. Dickinson, in his *Economic Motives*. I agree with Professor Dickinson that there may be special aspects or corners of economic activity upon which the researches of the psychologist may throw important light. On the other hand, there are important elements in economics which are not to be subsumed under "human behavior."

(In Professor Mitchell's 1910 papers his general allegation respecting hedonism was supplemented by somewhat similar specifications.) Now this is vastly better than tossing off Ricardo as a hedonist, but it is not enough.

It is impossible to deduce Ricardo's system from psychological postulates, unless the postulates are made so numerous that there is a one-to-one correspondence with every separate *element* in his system.[1] In short, Ricardo's postulates are found only *in* his system; in fact, I believe it substantially accurate to say that they *are* his system. Professor Mitchell has charged economists with having rationalized economic behavior. He himself, following, I admit, a goodly number of precedents, has rationalized economic science. For exposition's sake, or because they thought logical consistency required it, or for some much more subtle and obscure reason, economic theorists have often presented their doctrines as though they flowed from some first principle. But the first principle is generally purely ornamental,[2] like the meaningless "desire for wealth." The real soundness of a system of thought depends upon its internal consistency and upon the accuracy with which it summarizes the pertinent parts of experience. These considerations hold true not only for so-called deductive economics (the "deduction" is only a

[1] I cannot define an "element" except as that which would in strict logic demand a separate postulate. There are scores, possibly hundreds, of such elements in Ricardo's economics.

[2] When Professor Fetter substituted "choice" for utility as a first principle, the change not only did not but could not, of itself, affect his practical conclusions. Similarly, it is obvious that we could "base" Professor Mitchell's explanation of business cycles upon an explicit hedonism without altering his findings. The utility of fictional first principles is expounded in Vaihinger's philosophy of the "as if" (lucidly summarized by Havelock Ellis, in *The Dance of Life*, chap. III).

matter of expositional form [1]) but also for "institutional" or any other variety of economics.

Imagine a scholar from Mars, equipped with (1) a sufficient knowledge of our physical environment and its behavior, (2) a sufficient knowledge of the institutional relations (not behavior) of men, and (3) a "postulate," such as, say, that men behave as good hedonists should. Require him to formulate a political economy, and he would be helpless. Tell him then "men desire wealth," and he would reply, "But you have told me that wealth *is* whatever men desire that is not to be had freely." Give him Professor Mitchell's revised Ricardian postulates, and he would, for practical purposes, be equally helpless. But give him Ricardo's general fund of human experience, including his knowledge of the way men behave in their business relations and of the way prices, rents, and other things behave, give him also Ricardo's knowledge of what Smith, Malthus, West, and others had said, give him Ricardo's interest in certain problems of his day, give him finally Ricardo's power of generalization, and he could write Ricardo's *Principles*.

One cannot tell just how many "postulates" deductive economics requires. Nor would there be any practical purpose in trying to find out. The postulates change and grow as experience grows and as experience is better formulated. A very considerable part of our postulates can never be found anywhere else than in that direct and subtle kind of knowledge we call "common sense." [2]

[1] "The distinction between theory and practice we owe, indeed, like many another distinction, to the Greeks, but the divorce is our own contrivance. Accordingly, a theoretical knowledge of human, as of animal, behavior would mean to them the sort of knowledge that corresponds to the facts, and enabling one to deal with the facts in a practical way." — Lane Cooper, *The Greek Genius and Its Influence*, p. 10.

[2] Even in pure geometry, the task of accurately determining the neces-

Even a fuller knowledge of "economic motives" does not help much, for "motives" are merely names for general classes of behavior. I do not feel that Professor Douglas's interesting discussion of non-commercial incentives substantially increases our stock of useful knowledge. Professor Douglas had his rubrics ready-made. He looked for apt and interesting illustrations and found them. To be considered, also, is the notorious unreliability of the data available for such a study.[1] Nevertheless, studies like Professor Douglas's help to give a concrete content to our general notions.

The "economic man" was a fiction, not of economics, but of the sterile formal logic of economic method, for much of which economists themselves have been responsible. What the economists really abstracted was one special class of human relationships. Or better, they took into account the world as they knew it, modified only in that their eyes were focussed on the impersonal types of relationships that are to be found in the market places of countries where modern commercial institutions have been developed. They did not rationalize human behavior; they merely put into generalized form[2] their knowledge of some very real

sary postulates and fixing their minimum number has been difficult. And in general the task of putting "rigor" (strict deductive form) into pure mathematics has always lagged behind the growth of mathematical knowledge. In mechanics (a much simpler subject than economics) a rigidly deductive form is impossible. Not only inductive facts but common sense must be drawn upon. "Common sense, or better perhaps, good sense, comes close to being that *esprit de finesse* which Pascal, in a famous passage, contrasted with *l'esprit géométrique*." — Léon Lecornu, *La Mecanique*, p. 1.

[1] For example I do not doubt that Mr. Ford is perfectly sincere in saying that he has been motivated by "the ideal of service." But no one is a reliable judge of one's own purposes.

[2] Induction and deduction are terms properly applicable to forms of exposition and verification, not to thinking. I hazard a guess that in the actual construction of economic science analogies and "models" have

types of behavior. Their knowledge was incomplete, their generalizations were often too loose or too rigid, they failed to take account of all the available experience that was important for their purposes. But their method was not at fault, for, outside of the field where controlled experiments are possible, it is the only method science knows. A better knowledge of general psychology might have helped them here and there. It could not substantially have changed their postulates.

Professor Mitchell might conceivably say to all of this that he would consign the whole theory of value and distribution to the scrap heap, and that he would pin his faith to what he describes as "an unorthodox type of economic theory — a type of theory that deals with a range of problems undreamt of in the philosophy of value and distribution." This is "institutional economics." Its older forms, for one reason or another, were defective. "But by the eighteen nineties, Sidney Webb in England, Werner Sombart in Germany, and Thorstein Veblen in America were studying the evolution of economic institutions in a scientific, as opposed to an historical or a propagandist, spirit. Further, they were claiming that work of this kind is economic theory."

Consider the general character of Veblen's work. Sombart's is much like Veblen's, Sidney Webb's is substantially different. Veblen is a man of genius, but the

had much to do with the form generalizations have taken. Thus our generalizations relating to diminishing and marginal utility are not "deductions" from any psychological postulates, but appear to be organizations of our knowledge of a larger field of economic behavior upon the analogy of or after the model of generalizations already successfully established with respect to the behavior of the market place. This hypothesis would help to explain the fact that the notion of marginal utility served in its turn as the model upon which some able Austrian psychologists organized a yet wider body of experience into an interesting type of psychology.

term scientist does not fit him. He is something that may be as good or better: an artist, an impressionist, painting the picture of the world as he sees it. No one else would see it in just the same way, except through his eyes.[1] In modern stagecraft a slight difference in lighting changes the whole scene. So it is with institutional economics.

[1] Sometimes, however, writers of keen vision may see certain details alike. Says Joe Scott in Charlotte Brontë's *Shirley*, "Them that's mechanics, like me, is forced to think. Ye know, what wi' looking after machinery and sich like, I've getten into that way that when I see an effect, I look straight out for a cause." Cf. Veblen's collected economic essays, *passim*.

XIII

THE MEASUREMENT OF CHANGES OF THE GENERAL PRICE LEVEL[1]

THERE can be no final agreement upon any system of index numbers until there is agreement upon the precise meaning of the phenomena they are supposed to record. We must formulate our problem before we can hope to solve it. This paper is offered, therefore, as a preliminary exploration in the field of the general theory of index numbers of price movements. In this field there are two fundamental questions: (1) Just what are the measurable phenomena defined by the phrase, *changes of the general level of prices?* (2) Similarly, what is the precise meaning to be given to the phrase, *average changes of prices?* For the present I limit myself to the first of these two problems.

For convenience we may begin by taking the general level of prices as defined by P in the equation of exchange, $P=M/T$, where M is the amount, in money units, of money and of bank deposits exchanged during a year or other definite short period of time for the goods whose general price level is sought.[2] T is the quantity of goods exchanged for M, measured in the physical units to which price quotations refer. P is thus the (arithmetic) average amount of money paid by buyer to sellers per physical unit of goods purchased. The problem

[1] Reprinted, slightly abridged, from the *Quarterly Journal of Economics*, vol. XXXV, no. 4 (August, 1921).]

[2] M, as used here, corresponds to $(MV + M'V')$ of Professor Fisher's formulation of the equation of exchange.

is to determine a series of numbers whose variations shall have the closest possible correlation with the variations of P in successive periods of time.

Now a price, as a statistical datum, is *always a ratio*. A quantity of money is not a price, apart from its relation in exchange to a definite quantity of goods. Just as P, the average price of all goods bought and sold, is the ratio between M and T, so any price, p, is the ratio between m and t, where m is the number of money units paid for t units of goods.

Commonly, price quotations run in terms of the ratios between various amounts of money and single units of goods, that is, in terms of *unit prices*. The consequent of the price ratio is made unity, and price is expressed in terms of the antecedent alone. But this convenient practice should not be permitted to obscure the fact that unit prices are themselves ratios, that a price of two dollars per bushel, for example, may be expressed not only as the ratio between 2 and 1, but quite as accurately as the ratio between 4 and 2, or between 30 and 15, or as any other equal ratio.

A good example of the significance of the point that in devising index numbers prices must be treated as ratios is given by what may be termed the paradox of the arithmetic and harmonic means. "In European countries," says Professor A. C. Pigou,[1] "price is usually measured by naming the number of units of the standard of value which will buy a unit of the commodity; in India it is measured by naming the number of units of the commodity which can be purchased by a unit of currency. The choice between these two ratios is obviously a pure matter of arbitrary convention. But, to combine price ratios taken on the Indian plan into an arith-

[1] *Wealth and Welfare*, pp. 35, 36.

metic mean is equivalent to combining similar ratios taken on the European plan into a harmonic mean!"

By "price ratios," it seems from the context, Professor Pigou means relative prices — the ratios of prices in one year to the prices of the same goods in another year. Thus if the unweighted arithmetic mean of relative prices is $\Sigma\left(\frac{p_1}{p_0}\right)/n$, the reciprocal of the harmonic mean is $\Sigma\left(\frac{p_0}{p_1}\right)/n$, which is the arithmetic mean of relative prices expressed in the Indian manner, such as $\frac{1/p_1}{1/p_0}$.

The effect of using the reciprocal of the harmonic mean of relative prices instead of their arithmetic mean is to shift the base from one to the other of the two years involved in the comparison. That is, the reciprocal of the unweighted harmonic mean of relative prices is identical with the unweighted arithmetic mean, except that the base is shifted.

The significance of this lies in the fact that when basing years are shifted the results indicated by the use of the arithmetic average of relative prices are modified. Neither the harmonic mean of a group of ratios or its reciprocal will accord with their arithmetic mean. Yet prices quoted in the Indian manner are the reciprocals of prices quoted in the way to which we are accustomed. What is more important, the market facts are wholly independent of the manner of quotation. "Five cents a pound" and "twenty pounds for a dollar" are identical prices. Where the basic facts are the same the two

methods of quotation should lead to identical index numbers.

Similar considerations hold with respect to unweighted averages of simple or "absolute" prices. In general the arithmetic mean, $\Sigma(p)/n$, is not in agreement with the harmonic mean, $n/\Sigma\left(\frac{1}{p}\right)$. And no system of weighting, unless it be wholly arbitrary, will bring the two means into agreement, so long as the prices averaged are treated as quantities instead of ratios. To treat p as merely a sum of money, five cents for example, and its reciprocal $1/p$ as a quantity of a commodity, twenty pounds for example, is to handicap ourselves by a serious initial error.

But if we hold rigorously to the fact that price is always a ratio, involving two magnitudes — a quantity of money *and* a quantity of goods — the problem may be said to solve itself. The way in which prices happen to be quoted does not affect the result. Following the notation already indicated, put m/t for p. Then the only arithmetic mean which has any obvious meaning is the weighted mean, $\Sigma(m)/\Sigma(t)$, or $\Sigma(tp)/\Sigma(t)$. In a similar way the weighted harmonic mean is $\Sigma(m)/\Sigma(m/p)$, or $\Sigma(tp)/\Sigma(t)$. It is identical with the weighted arithmetic mean. This weighting of the arithmetic mean of actual prices by physical quantities of goods and of the harmonic mean by quantities of money, or "values," is in no sense arbitrary. It is clearly indicated or even compelled by elementary considerations.

The paradox which has just been discussed, like a host of other difficulties which beset the construction and interpretation of index numbers, springs from erro-

neous methods of handling ratios. These errors have crept into the treatment of the problem, I believe, through the habit of identifying ratios with the fractions commonly used to express them. Fractions whose numerators are the antecedents of ratios and whose denominators are their consequents may be subjected to certain mathematical operations without affecting the relations they express. But addition and subtraction are not among these legitimate operations.

Moreover, when the fractions used to express ratios are added as part of the process of finding their arithmetic mean, they are first reduced to a common denominator. This involves an arbitrary weighting of the numerators. In common practice the numerators are determined by the condition that the denominator shall in each case be unity. Then the denominators are forgotten, the numerators are termed "prices," and their mean is held to be an "average price." What the unweighted arithmetic mean of actual prices really gives is merely the average money cost per unit of a bill of goods consisting of one unit each of the different commodities whose price quotations are used. Similarly the reciprocal of the unweighted harmonic mean gives the average number of physical units of goods exchanged for one unit of money. The difference between the results given by the two means is due to the presence in each case of weighting that is none the less arbitrary because it is not intended. In one case single units of goods and their money costs are the components of the average. In the other case its components are units of money (dollars) and "dollar's worths" of goods. One mean has as good a claim to represent the general price level as the other. Each represents a

group of prices whose peculiar composition is dictated by the method by which the mean is constructed.

We have seen, however, that with *complete* weighting the two means lead to consistent results. The reciprocal of the arithmetic mean becomes identical with the arithmetic mean of the reciprocals of the components of the original mean. This illustrates an elementary but important difference between the unweighted and the completely weighted arithmetic mean of ratios. The unweighted arithmetic average of ratios is a "fictitious mean" of the familiar type, but its claim to represent or typify a group of ratios is marred by its hidden weighting and especially by the fact that it gives different results according as the ratios whose mean it purports to be are expressed by one form of fraction or another. But the completely weighted arithmetic average is not a "fictitious mean." In fact its title to the name average or mean might even be disputed. It is constructed by a process of summation rather than of "averaging." Its antecedent and consequent are the respective sums of the antecedents and consequents of its constituent ratios.

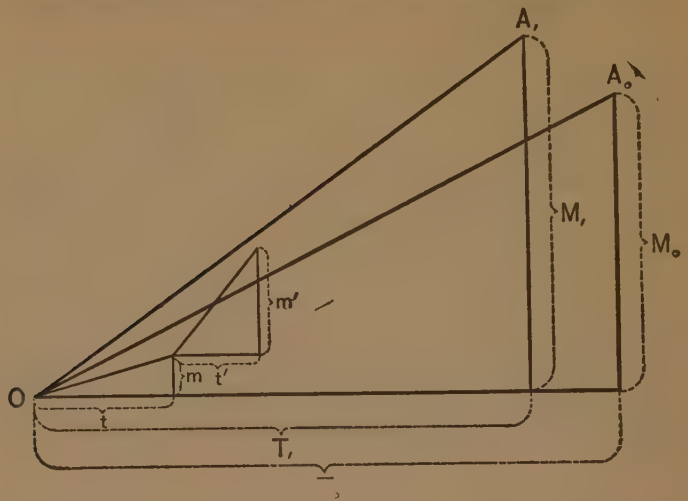
Consider an example taken from another statistical field. The proportion of the population of each State found in some specified population group, those for example who are gainfully employed, is reported by the federal census. For any one State this proportion is the reciprocal of the ratio of the whole population of the State to the number who are gainfully employed. An unweighted mean of the per cents which express the proportions found in the different States might conceivably be held to represent the proportion gainfully employed in an imaginary "average State." But its reciprocal is likely to differ, perhaps rather widely,

from the unweighted mean of fractions which have whole populations as numerators and numbers gainfully employed as denominators. Yet this second mean has as valid a title as the first to be the ratio attributed to the "average State."[1] With complete weighting the two means are brought into agreement. But they no longer indicate a "fictitious," "representative," or "typical" condition. They express for the United States as a whole the ratio between the aggregate number gainfully employed and the aggregate population. In a similar way the completely weighted arithmetic mean which we have been discussing is not, properly speaking, an average of separate classes of prices. The ratio between the total amount of money exchanged for goods and the total quantity of goods exchanged for money directly expresses the general or aggregate price level.

Some aspects of the matter may be brought into a clearer light by a geometrical device. The equation of exchange may be represented by a right-angled triangle,

[1] A somewhat similar problem is created by the requirement of the Constitution that "Representatives shall be apportioned among the several States according to their respective numbers." The arithmetic mean will in general lead to different apportionments according as "representatives" or "respective numbers" are denoted by the numerators of the fractions used to express apportionment ratios. Precedent and tradition support the present practice of using the first of these two forms. This gives results which in the long run are more favorable to the larger States than would be obtained by using the reciprocals of such fractions. Dr. J. A. Hill and Professor E. V. Huntington have shown that this particular problem calls for the use of the geometric rather than the arithmetic mean. With the geometric mean, of course, the form the fractions take is a matter of indifference, for the geometric mean of the reciprocals of fractions is the reciprocal of the geometric mean of the fractions in their original form. [Cf. the "Report upon the Apportionment of Representatives," by the Census Advisory Committee, in *Quarterly Publications of the American Statistical Association*, December, 1921.]

as in the accompanying diagram. The horizontal and vertical sides represent T and M, respectively, while P is the tangent M/T, the slope of the third side, OA.[1]



Knowledge of the absolute magnitudes of M and T would give us the value of P. But the most available and certainly the most trustworthy data are observations on the price ratios at which different goods are exchanged for money. These may be supplemented, we must assume, by some knowledge of the relative magnitudes of the volume of trade in the same goods. That is, our data are certain p's and t's. Our facts are incomplete.

What is the best method of inferring from our partial

[1] Similarly, the tangent M_1/T_1 is P_1, the price level in some other year. Individual weighted prices are represented by m/t, m'/t', etc. They are placed on the diagram so as to illustrate the summational method, viewed as the addition of vectors. To illustrate the method of least squares these small triangles should be placed so as to have a common origin at O.

observations the most probable value of P, that is, the most probable slope of the line OA?

If each reported price (itself to be interpreted graphically as a tangent or slope) be taken as a separate and independent observation on the value of P, that is, on the tangent or slope M/T, the problem is obviously that of fitting a straight line to a number of observed points, each subject to unknown but presumably unbiased errors. The method of least squares suggests itself. This method, it will be remembered, involves the principle of the arithmetic mean. The appropriate formula is:

$$P = \Sigma(tm)/\Sigma(t^2) = \Sigma(t^2p)/\Sigma(t^2).$$

Where only one physical unit of each commodity is taken into account this formula becomes $\Sigma(p)/n$, the unweighted arithmetic mean of unit prices. Graphically this may be interpreted by conceiving the slope of the line as determined by passing it through the origin and a point fixed by the arithmetic mean of the n values of the ordinate m at the point where the abscissa t is unity.

Taking the different magnitudes of t into account, the method gives a less familiar result. The ratio $\Sigma(t^2p)/\Sigma(t^2)$ is an arithmetic mean weighted in accordance with the squares of the numbers of commodity units sold at the stated prices. That is, it is what is commonly deemed a properly weighted arithmetic mean, operated upon by again multiplying its different component terms by the weights already used.

But the old difficulty remains. Invert the triangle, using its base as an axis, and then rotate it in a clockwise direction so that its vertical side becomes horizontal. The slope of the hypothenuse is now T/M, or $1/P$. The formula for determining this slope, accord-

ing to the method of least squares, is $\Sigma(mt)/\Sigma(m^2)$, or $\Sigma\left(\frac{m^2}{p}\right)/\Sigma(m^2)$. Letting m equal unity this gives $\Sigma\left(\frac{1}{p}\right)/n$, the reciprocal of the unweighted harmonic, not of the unweighted arithmetic mean. Nor is the weighted form, $\Sigma(mt)/\Sigma(m^2)$, the reciprocal of the correlative form previously found, $\Sigma(tm)/\Sigma(t^2)$. Their product is $[\Sigma(mt)]^2/[\Sigma(m^2) \times \Sigma(t^2)]$, which is not, in general, equal to unity.

As between these two weighted forms there seems to be no ground of preference. But they may be combined or averaged so as to give each of them equal weight. This is best done by using the geometric mean. The result is $\sqrt{\Sigma(m^2)/\Sigma(t^2)}$, or $\sqrt{\Sigma(t^2p^2)/\Sigma(t^2)}$. The method of least squares, therefore, does not lead to the ratio of aggregates, but to another weighted arithmetic mean, the square root of a weighted average of squares in which the weights are also squared.

The use of this formula, novel as its application to price statistics is, might perhaps be justified in cases where the price quotations available could properly be regarded as merely a relatively small random sample of the complete series of actual prices.[1]

In general, however, the analogy between the relation of observed prices to the general price level and the relation of points given by experimentation or observation to the equation of the best-fitting curve or line may easily be overstressed. Price statistics are not, in a strict sense, independent observations on an unknown

[1] [Professor Irving Fisher has tested this formula (*The Making of Index Numbers*, p. 380), using the prices of 36 commodities in 1913 and 1916, and finds it even less reliable than I should have expected.]

magnitude — the general price level. Each recorded observation is a partial rather than an independent report on that magnitude. The antecedent and the consequent of the reported price ratio enter into the sums which constitute the antecedent and consequent of the general price ratio.

If all exchanges were reported, the equation of exchange could be written $P=(\Sigma m)/\Sigma(t)$. And if, though not complete, our list of reported prices is reasonably large and representative, it may be used to give a *partial sum*. That is, we take $\Sigma(m)/\Sigma(t)$ as giving a workable approximation to M/T. We must assume, therefore, either that the remainders $M-\Sigma(m)$ and $T-\Sigma(t)$ are relatively small as compared with $\Sigma(m)$ and $\Sigma(t)$, or that the unknown ratio $[M-\Sigma(m)]/[T-\Sigma(t)]$ is not likely to be greatly at variance with the ratio $\Sigma(m)/\Sigma(t)$, and as likely to be above it as below it.

Graphically, this amounts to determining the slope of the line OA by passing it through a point whose ordinate is $\Sigma(m)$ and whose abscissa is $\Sigma(t)$. This is equivalent, of course, to the addition of a series of vectors, each of which is determined by a particular weighted price ratio, m/t, just as OA is determined by M/T.

With the general price level determined there is no technical difficulty in the way of comparing one year's prices with another's by constructing a series of index numbers representing the fluctuations of the general price level. It is necessary only that the price level in each successive year be expressed as an independent magnitude. With this condition met relative changes in the price level are easily determined. The results

do not depend upon the selection of a particular basing year or period.

With constant weighting $\Sigma(t)$ is constant, and the index numbers express merely the fluctuations of $\Sigma(m)$. In this form index numbers of the general price level are the "weighted sums" or "aggregative" index numbers which several recent writers have ranked among the better index numbers, and which have proved themselves serviceable in practice. They indicate the varying amounts of money payments required by sales of constant quantities of goods at prices determined in part by the condition that such sales have not in fact been constant. Or with other weights than t, they may indicate the changing market value of a constant volume of production, or the changing cost of a fixed bill of goods or of a fixed standard of living.

As compared with the weighted sum, the completely weighted average of actual prices has not met with equal favor. This may be accounted for by a variety of circumstances. Doubtless the limitations in the meanings of other types of index numbers, including weighted sums, have not always been fully realized. Moreover, the magnitudes, even the relative magnitudes, of t (as of other possible weights that might represent physical quantities) are so incompletely and inaccurately known as to discourage attempts to estimate and utilize their fluctuations in index numbers of the general level of prices. Most of all, I suspect, it has been felt that an average into which enter such heterogeneous things as tons, yards, gallons, and dozens, is devoid of meaning.

With respect to this last objection it should be observed that for ascertaining the price level there is no advantage in converting these incongruous physical units into some common measure, pounds for example.

This merely introduces a new and deceptive system of weighting. Moreover, price *ratios* are all that are involved in the form of weighted average which I have been discussing. The price ratio itself is taken as the fundamental statistical datum. In that ratio each physical unit figures merely as the quantum in terms of which prices are made and recorded.

A serious difficulty appears, however, when the changing volume of trade in different types of commodities is taken into account. Other things being equal, if sales of commodities whose unit prices are low come to make a larger proportion of the total volume of exchanges, the fact will be reflected in a lower general level of prices. Similarly, if the sales of commodities whose unit prices are high grow faster than other sales, the general level of prices will be higher.

For this reason, and, so far as I see now, for this reason only, trade should be measured in "dollar's worths" rather than in single physical units. Conversion of physical units into "dollar's worths" may be effected on the basis of the prices that prevail in either of any two years whose price levels are to be compared. For that year the general price level becomes unity. For the other year it becomes $\Sigma(m_1)/\Sigma(t_1p_0)$, or $\Sigma(t_1p_1)/\Sigma(t_1p_0)$, where t still represents the number of physical units sold at the unit price p, and the subscripts 0 and 1 indicate respectively the basing year and the other year.

This result, $\Sigma(t_1p_1)/\Sigma(t_1p_0)$, may be taken as an expression of the *relative price level.* Like the absolute price level it is a ratio between the volume of money payments and the quantity of goods purchased, but in this case the quantity of goods is measured not in

tons, yards, bushels, etc., but in the amounts that were exchanged for a dollar in the basing year. In form, however, this result is merely a ratio between weighted sums of unit prices, the constant weights, it will be noted, being not those of the base year, but those of the year whose price level, relative to that of the basing year, is sought.

One difficulty remains. If the basing year is shifted, so that the price level in what had been the basing year is now expressed as relative to that of the other year (which now becomes unity), we have the form, $\Sigma(t_0p_0)/\Sigma(t_0p_1)$. This is not in general the reciprocal of the other form, $\Sigma(t_1p_1)/\Sigma(t_1p_0)$. Here again the geometric mean may properly be used to combine the two inconsistent results.[1] We then have as the expression of the relative general price level in the year denoted by the subscript 1: $\sqrt{\frac{\Sigma(t_1p_1)}{\Sigma(t_1p_0)} \times \frac{\Sigma(t_0p_1)}{\Sigma(t_0p_0)}}$.

I have the more confidence that this is an acceptable solution of the particular problem in hand because Professor Irving Fisher has already reached precisely this result and has adjudged it to be the best of the different formulas he has examined and tested.[2] Furthermore, Professor Fisher's result has the weighty approval of Mr. C. M. Walsh.[3] So far as I know, however, the analysis by which I have reached this conclusion runs along lines which differ from those Professor Fisher and

[1] The propriety of using the geometric mean hinges on the fact that if the two expressions were in agreement their *product* would be unity.

[2] *Quarterly Publications of the American Statistical Association*, March, 1921, p. 536. [Professor Fisher, however, prefers to view the formula as giving a measure of average changes in prices, not of changes of average prices.]

[3] *Ibid.*, p. 539.

Mr. Walsh have followed. It is possible, therefore, that in attempting to find my way into the fundamentals of the problem I have done something to strengthen the logical foundations of Professor Fisher's index number.[1]

On the basis of the evidence now in hand I believe this to be the best single index number of the general level of prices. It is less likely to be deceptive than any other formula that I happen to know. I mean that it gives a more direct and unequivocal answer to the particular questions most students of changes of the price level are likely to have in mind. But I fear that no single index number will afford a sufficient answer to all such questions.

It should be clear, moreover, that in the present paper we have been concerned solely with the problem of changes of the general price level. The measurement of average changes in prices is in some respects a distinct and different problem.[2]

[1] It is proper to say that most of the results of the present paper were presented to the Economics Seminar at Cornell University in 1919. I did not suggest at that time, however, that the geometric mean be used to effect a compromise between two inconsistent results.

[2] But see below, p. 294.]

XIV

FISHER'S "THE MAKING OF INDEX NUMBERS" [1]

In an appendix to this volume[2] Professor Fisher gives a brief list of "landmarks in the history of index numbers." In any future list, it is safe to say, an important place must be assigned to Professor Fisher's own work. Using new methods, he has pushed his explorations in certain directions further than any of his predecessors. He has surveyed and mapped new territories and has definitely conquered no small part of them. Even if he has not reached his objective he has come within sight of it. Debatable territory — a no man's land as yet — remains. But its area is relatively small, its boundaries are known, and the general nature of its difficult terrain is clearly indicated. In short, Professor Fisher's work on index numbers is a notable scientific achievement.

The more important studies of index numbers heretofore have generally been either experimental or analytical.[3] Not the least of the merits of Professor Fisher's work is that he avails himself of both methods, even though he leans rather more to the empirical than to the analytical side. He proceeds by setting up several

[1 Reprinted, slightly abridged, from the *Quarterly Journal of Economics*, vol. xxxvii, no. 2 (February, 1923). Some minor modifications have been made in the light of Professor Fisher's comments in the same journal, vol. xxxvii, no. 4 (August, 1923).]

[2] Irving Fisher, *The Making of Index Numbers. A Study of their Varieties, Tests, and Reliability.* Boston and New York, Houghton Mifflin Company. 1922. Pp. xxxi, 526. (Publications of the Pollak Foundation for Economic Research, No. 1.)

[3] It is proper to say here that Professor Fisher's work, like other recent studies of index number construction, serves to make clear the fundamental soundness of Mr. C. M. Walsh's general analysis of the problem in his work, *The Measurement of General Exchange Value* (1901).

tests of the accuracy of index numbers and then by passing a long series of different types of index numbers under review, ranking them finally in accordance with their degrees of conformity with the stated tests, as well as by the ease by which they may be calculated and by other practical considerations.

Altogether Professor Fisher tests 134 different formulas, all variants, as he suggests, of six fundamental types: the arithmetic, harmonic, geometric, median, mode, and aggregative (or weighted sum). Included in this long list are substantially all the forms of index numbers that have ever been used in practice or suggested by writers on the subject. Many of the forms are novel, but most of these are developments of familiar formulas, obtained by manipulating basing years or weights. Some (not all) of these new forms — and some of the old, for that matter — are merely fanciful. The purpose, it is clear, was to make the list absolutely complete. But it is impossible for any one really to exhaust all of the feasible forms of index numbers. In Professor Fisher's list, in fact, there are some fairly obvious omissions. Most of these are to be attributed to his preference for one type of weights, namely, sums of money or "values" (in conventional notation, $q \times p$).

For his tests Professor Fisher uses the prices of 36 commodities in the six years from 1913 to 1918. These price series are selected from the 1474 collected by Professor Wesley Mitchell for the War Industries Board. They are admirably adapted to the purpose because, first, the data include quantities as well as prices, and, second, because, as Professor Fisher suggests, the period covered was one of "extraordinary dispersion in the movements both of prices and quanti-

ties." Such data afford a severe test of the reliability of index numbers. It is therefore highly significant that the results given by the better formulas agree very closely. By about forty different formulas prices in 1918 were found to be between 77.3 and 78.2 per cent higher than prices in 1913. The results of the eleven formulas Professor Fisher deems best have the narrow range of between 77.63 and 77.83 per cent, and five of these give between 77.63 and 77.67 per cent. By less reliable formulas the increase in prices is reported to be as little as 67 per cent or as much as 144 per cent. Some difference of opinion may remain respecting just what, for a particular purpose, is the one best formula, for it appears that there are a number which are almost equally good. But there can be no longer any doubt that some formulas — including a few that have been commonly used — are untrustworthy. Such are the positive and extremely valuable results of Professor Fisher's work.

The practical significance of this general winnowing of the whole field of index numbers can hardly be overestimated. In the past the largest errors of index numbers have been put into them by improper methods of construction. Errors in the data have by comparison been negligible, and even errors introduced by biased selections of data have generally been of minor consequence. But now that Professor Fisher has shown that the index number may be an instrument of precision, that the instrumental error of the formula may be reduced to a small fraction of 1 per cent, no excuse remains for the use of any but the most accurate formulas. Just how accurate our index numbers are should depend almost wholly upon the accuracy and the representative character of the data we use.

Before passing to Professor Fisher's criteria of a good index number it will be well to consider a criterion he rejects — the so-called "circular test." This test was suggested by Professor H. L. Westergaard, given much weight by Mr. C. M. Walsh, and accepted by Professor Fisher himself in his earlier work, *The Purchasing Power of Money*. By this test, index numbers should be *self-consistent*, that is, they should be independent of any particular base. Chain and fixed-base index numbers should give identical results. In general, if I_{xy} is the index number for the year or place x as compared with the year of place y, the following relation should hold:

$$I_{ac} = I_{ab} \times I_{bc}$$

Now I go so far with Professor Fisher as to agree that among the best index numbers are some that do not meet the circular test. That test is, in principle, inconsistent with *some* purposes for which index numbers may properly be constructed. But his conclusion that "the circular test is theoretically a mistaken one" is too sweeping. His point is that the weightings which may be appropriate as between any two years or places may not be appropriate as between either one of them and a third year or place. "If we are intent on getting the very best comparison between Norway and Egypt, we shall not go to Georgia for our weights. In the direct comparison between Norway and Egypt the weighting is, so to speak, none of Georgia's business."

This argument would be convincing if index numbers were used *only* for year-with-year or place-to-place comparisons. The fact is that for most purposes we need *series* of index numbers. First, because it is impracticable to compute separate index numbers for all possible combinations of years. Second, because the

successive changes of the series, its fluctuations and trend, may be the principal objects of our interest. Surely index numbers have other purposes than comparing the price level of to-day with some *one* earlier year. The circular test must be insisted upon as an important criterion of a good *series* of index numbers.

In practice an index number has to serve both for year-with-year comparisons and as a member of a series. Our choice of formula must therefore depend upon (1) the relative importance of self-consistency in the series and of accuracy of comparison with some *one* selected year, and (2) the *degree* in which the formulas best by one test are inaccurate by the other. The second of these two considerations does not point definitely in one direction or the other. It is impossible to lay down as a general rule that the year-with-year comparisons afforded by the best of the index numbers that meet the circular test are either more or less accurate than the general comparisons-in-series derived from such index numbers as best express the change of prices as between any one year and the basing year. Much depends upon the length and general character of the period covered by the series. One thing, however, is clear. For short periods the errors of the best index numbers, one way or the other, are exceedingly small.

I agree with Professor Fisher respecting the formulas which give the best year-with-year comparisons. But I believe that in the practical construction of standard series of index numbers it is quite as important, on the whole, that the series should be self-consistent as that each number of the series should afford an impeccably accurate comparison with the base of the series.

In practice the choice resolves itself into one between constant and what Professor Fisher calls "crossed"

weights. Simplicity, ease of computation, and the availability of the necessary data are all on the side of the constant-weighted index numbers which conform to the circular test. Moreover, over short periods of time they do not differ greatly from cross-weighted or crossed-formula index numbers. Thus, using the ordinary aggregative with constant basing-year weights ($\Sigma q_{13}p_{18}/\Sigma q_{13}p_{13}$), the index number for 1918 for the prices of Professor Fisher's 36 commodities, on a 1913 base, is 177.87. Using quantities in 1918 as weights makes the index number ($\Sigma q_{18}p_{18}/\Sigma q_{18}p_{13}$) 177.43. Professor Fisher's "ideal" crossed-formula index number is the geometric mean of the two, 177.65. For practical purposes these differences are negligible.

A properly weighted geometric average gives results substantially as good. Professor Fisher, like every one else, weights the geometric, by "values." This introduces an element of biased error. He tests four different sets of constant weights, viz., $q_{13}p_{13}$, $q_{18}p_{13}$, $q_{13}p_{18}$, $q_{18}p_{18}$. The first two forms, in which prices in 1913 are elements in the weights, give results that as compared with the "ideal" are about five points too low. The other forms, in which prices in 1918 are elements in the weighting, give results that are about five points too high. The explanation is simple. In the first two forms weights and increases of prices are inversely correlated. In the other two forms the correlation is direct. Using arithmetic means of "values" in 1913 and 1918 as weights lessens, but, Professor Fisher thinks, does not eliminate the bias he finds in the weighted geometric, for the index number thus obtained for 1918 is 179.54, too high by about two points.[1] But when he uses

[1] Professor Warren M. Persons has shown empirically (in the *Review of Economic Statistics*, May, 1921) that a geometric average with con-

geometric rather than arithmetic means in computing his weights he removes, it appears, substantially all bias.

Further evidence of the reliability of properly constructed index numbers with constant weights is afforded by Professor Fisher's test of the Bureau of Labor Statistics index number of wholesale prices. He finds that his ideal formula, using the Bureau's price quotations and weights, gives an index number for the year 1919 on a 1909 base that is 1.4 per cent lower than the Bureau's index number with weights computed in accordance with data for 1909, and that much higher than the Bureau's new index number in which the weighting is in accordance with data for 1919. "The adjustments needed for the intervening nine years," says Professor Fisher, "barely exceed 1 per cent in any case." These results should increase the confidence students already have in the Bureau's index number. They indicate that when simplicity and conformity to the circular test are taken into account, standard series of index numbers are best constructed as aggregatives or geometric averages with constant weights. The Bureau's practice of revising its constant weights once in ten years seems adequately to meet all reasonable requirements. For special studies of the movement of prices or of the cost of living between any two dates the crossed aggregative type is unquestionably superior. Beyond much doubt it gives the most accurate expression of the

stant (broadened-base) weights may agree very closely with the "ideal" crossed aggregative type Professor Fisher prefers. Professor Persons observes that his investigations "point to the conclusion that the differences between Fisher's indices and the geometric average with constant weights result primarily from differences in weighting." Further, "Fisher's index is 'the best' measure of general exchange value if we desire to compare two years only, and not a series of years each with the other."

difference between the general price levels that prevail in any two years.

Professor Fisher's own criteria of the quality of index numbers are what he calls the factor-reversal and time-reversal tests. The first of these is new. It appears to be prompted by Professor Fisher's special interest in monetary problems. In a general movement of "values," i.e., of total money payments, changes both in prices and in the output of goods or the volume of trade are at work. It is desirable to disentangle the effects of the price factor. For this purpose, we need index numbers of prices which, combined with similarly constructed[1] index numbers of quantities, will agree with index numbers of "values," that is, of prices multiplied by quantities produced or exchanged. If P is an index number of prices, Q of physical quantities, and V of values, the condition, algebraically expressed, is:

$$P_{ab} \times Q_{ab} = V_{ab} = \frac{\Sigma q_b p_b}{\Sigma q_a p_a}$$

The idea is ingenious and important. As Professor Fisher suggests, the principle involved is recognized when bank clearings or exports and imports are divided by an index number of prices in order to get a rough index of the volume of domestic or foreign trade. Ability to meet the test, it must be granted, is a quality highly desirable in an index number. But inability to meet the test *algebraically* is not, I believe, sufficient ground for discarding a formula. Probabilities as well as algebra must be taken into account. For example, it can be shown analytically that there is a high degree

[1] The only difference between the index numbers of prices and of quantities is that the p's and q's are interchanged.

of *probability* that a properly weighted geometric average will agree very closely with Professor Fisher's ideal formula. And Professor Fisher's own tests show a number of cases of strikingly close agreement between index numbers which do and others which do not rigidly satisfy the test. In practice the character of the available materials may properly determine the structure of an index number.[1] Thus where weights are not available it may be best to use the geometric average, despite the fact that *no* unweighted average meets the factor-reversal test. But the suggestion of this test is undoubtedly an important contribution to the theory of index numbers.

The time-reversal test is not altogether new, although its importance has too often been forgotten. It is that an index number should conform to the condition,

$$I_{ab} \times I_{ba} = 1$$

That is, if an index number shows that prices were twice as high in a given year as in another, the same formula should show that prices in that other year were half as high as in the given year. All index numbers that meet the circular test conform likewise to the time-reversal test, but not all index numbers that conform to the time-reversal test will meet the circular test. In short, the time-reversal test is a part of the general circular test. Professor Fisher rejects the whole but makes the part the cornerstone of his structure of index-number theory. There is nothing illogical in this. It is consistent with the emphasis he puts on securing the utmost accuracy of the statement of the movement of prices between *pairs* of years.

[1] Professor E. E. Day informs me that his use of an average of relatives in his index numbers of physical production was determined in part by the character of the data available as weights.

Closely and logically connected with the tests he prefers is Professor Fisher's discussion of bias — one of the most valuable, as well as most original, parts of his work. In recent years students of index numbers have become aware of the systematic bias inherent in certain types of index numbers,[1] especially the arithmetic and harmonic averages of relative prices. Professor Fisher is the first to give the problem the emphasis and the systematic treatment it deserves.

Index numbers, Professor Fisher holds, are affected by two sorts of bias: type bias and weight bias. Both sorts of bias may be revealed and measured by the time-reversal test. The amount of bias in an index number that compares prices of one year with those of another is half the difference between the results obtained by taking each year in turn as the base. More accurately, the relative bias is revealed by the percentage difference between each of the two results and their geometric mean. The forms most seriously affected by type bias are arithmetic and harmonic averages of relative prices. Those most affected by weight bias are geometric averages.

Weight bias, as I have suggested, is easily explained. When the weights are "values," which contain as elements prices of the earlier of the two years compared, rising prices are relatively underweighted and falling prices relatively overweighted. When prices of the later year are elements in the weights, rising prices are relatively overweighted and falling prices relatively

[1] Professor Fisher refers to Mr. C. M. Walsh's early recognition of the existence of bias (1901). He fails to note the fundamental importance of two early papers by Mr. A. W. Flux (*Memoirs and Proceedings of the Manchester Literary and Philosophical Society*, vol. XLI, Part III, Session 1896–97; *Quarterly Journal of Economics*, vol. XXI, pp. 613–31, Aug., 1907).

underweighted. Weight bias, it is important to observe, depends solely upon the character of the weighting and not upon the general type of the formula used for the index number. The geometric average is peculiarly sensitive to biased weighting, for the reason that its weights are exponents rather than coefficients.

There are some interesting points in the effects that biased weighting has upon arithmetic averages of relative prices. In Professor Fisher's notation the unweighted form is $\Sigma p_1/p_0 \div n$. Consider the effect of weighting by four different sets of "values," namely: q_0p_0, q_1p_0, q_0p_1, and q_1p_1. Using the first set of weights converts the index number into the form $\Sigma q_0p_1/\Sigma q_0p_0$. The second set gives $\Sigma q_1p_1/\Sigma q_1p_0$. These are the two fundamental aggregative formulas. They may be interpreted as ratios of sums of unit prices weighted by *quantities*. As we have already seen, they are among the more trustworthy index numbers. The bias in their weighting is small. The canceling of the p_0's in the weights and in the price relatives eliminates both type bias and weight bias. Except for this fortunate algebraic conjuncture, "values" are no more imperatively indicated as weights of arithmetic than of geometric averages of relative prices. This is shown by the effects of using the other two sets of weights. The weighted index numbers are

$$\frac{\Sigma q_0p_1\frac{p_1}{p_0}}{\Sigma q_0p_1} \text{ and } \frac{\Sigma q_1p_1\frac{p_1}{p_0}}{\Sigma q_1p_1}$$

The second is the form employed by Palgrave in a revision of the *Economist's* figures.[1] Both formulas are unreliable, giving results which, by Professor Fisher's

[1] Cf. Walsh, *Measurement of General Exchange Value*, pp. 538, 565.

tests, are (for 1918 compared with 1913) about 10 points too high. The upward bias of the weighting, born of the direct correlation of the weights and the price relatives, reinforces the inherent upward bias of the arithmetic average of price relatives.

Professor Fisher assigns the two aggregative forms given above to positions 36 and 38 in the "order of merit" of the 134 formulas he tests. If he had attached importance to the circular test, he would have been compelled to rank them closer to the top. But to the two other weighted arithmetic averages he assigns positions 104 and 106. It is clear that the ordinary arithmetic average of price relatives is a good index number only when it is so weighted that it is no longer an average of price relatives. The harmonic average of relative prices is affected by weighting in precisely the same way as the arithmetic. For it, however, "values" containing p_1's are the appropriate weights, for the harmonic is identical with the arithmetic with the basing year reversed.

Type bias is not quite so simple as weight bias. Despite the generally illuminating character of Professor Fisher's study of type bias, his *explanation* of it seems to me to be the least satisfactory feature of his work. He attributes type bias to the conjunction of two factors: (1) bias inherent in the arithmetic and harmonic averages, and (2) the amount of the dispersion of price relatives. In my opinion the properties of the arithmetic and harmonic averages have little or nothing to do with the matter. The dispersion of price relatives is at fault, but if it were not for the skewness or asymmetry of the distribution, the amount of the dispersion would not matter.[1] The asymmetry of the dis-

[1] Professor Fisher holds (p. 408) that "In choosing the formula for an

tribution, however, increases with the amount of the dispersion.

The property of the arithmetic mean which Professor Fisher finds at fault is its failure to equal the reciprocal of the arithmetic average of the reciprocals of its components. Put algebraically,

$$\frac{\Sigma x}{n} \times \frac{\Sigma \frac{1}{x}}{n} > 1.$$

Inverting the terms we get:

$$\frac{n}{\Sigma \frac{1}{x}} \times \frac{n}{\Sigma x} < 1,$$

which indicates the similar failure of the harmonic mean to live up to all the responsibilities Professor Fisher would like to put upon it. It follows that neither the unweighted arithmetic average or the unweighted harmonic average of price relatives will meet the time-reversal test. The arithmetic will have an upward and the harmonic a downward bias. But — and this is the important point — the fault must be imputed, not to the averages, but to the use of price relatives. The inequalities noted above do not, of themselves, create bias. Index numbers that completely conform to the time-reversal test may be constructed by using arithmetic or harmonic averages.

Take as components, for example, measurements of the statures of the boys entering a preparatory school.

index number the skewness or asymmetry of distribution is of absolutely no consequence." The context carries the inference that he means merely that the distortion due to skewness can be got rid of by taking the mean between two averages that are distorted in opposite directions, as, for example, the arithmetic and harmonic.

Arithmetic or harmonic averages of the measurements made in successive years will be series of index numbers of the general changes of the statures of the entering classes. The harmonic averages will be consistently smaller than the arithmetic. But each series will be consistent within itself, complying rigidly not only with the time-reversal but with the complete circular test. This fact is wholly independent of the degree or the character of the dispersion of the measurements. If the components are measurements of the same group of boys in successive years, so that a definite growth element is present, the index numbers will still remain self-consistent. And such will also be true of index numbers made up of unweighted arithmetic or harmonic averages of actual unit prices. Unweighted averages of actual prices are unreliable, but they have no "type bias."

The bias of the arithmetic and harmonic averages is to be attributed, therefore, not to the operations indicated by the forms of those averages, but to operating in such fashion *upon price relatives*. Right here is an elementary but fundamentally important point to which Professor Fisher has given inadequate attention. Price relatives, such as p_1/p_0, he refers to sometimes as "ratios," sometimes as "percentages." "An index number of prices," Professor Fisher holds, "shows the average percentage change of prices from one point of time to another." Or again, "An index number of the prices of a number of commodities is an average of their price relatives." Now there is a sense in which most index numbers, including the aggregative forms,[1] are

[1] That the ratio of aggregates, $\frac{a+b+c+\ldots\ etc.}{a_1+b_1+c_1+\ldots\ etc.}$, is a mean of (*moyenne entre*) the terms $\frac{a}{a_1}, \frac{b}{b_1}, \frac{c}{c_1}$ etc., is a theorem that dates back to Cauchy. Professor Fisher proves that it is, moreover, a true average.

"averages of ratios"; but no good index number is definitely an average of percentages, that is, of ordinary price relatives.

Ratios, as such, are not additive. A percentage, or price relative, is only one of an indefinite number of arbitrary forms that a ratio may take. Adding price relatives requires that the fraction which expresses a ratio be determined by the arbitrary condition that its denominator shall equal unity. The numerator is determined in part by this arbitrary condition and in part by the true magnitude of the ratio. Let $p_1 - p_0 = \Delta p_0$; $p_1' - p_0' = \Delta p_0'$, etc., where Δp_0, $\Delta p_0'$, etc., may be either positive or negative. Then the unweighted arithmetic average of price relatives takes the form

$$1 + \frac{\frac{1}{p}\Delta p_0 + \frac{1}{p_0'}\Delta p_0' + \cdots}{\square}$$

The weight given the amount of increase or decrease of any price is inversely proportional to the magnitude of that price in the base year. Such weighting brings the different amounts of change *toward*, but not *to*, a comparable basis. As soon as prices begin to move away from the base in dispersed fashion the proportions they originally bore one to another cease to obtain among them or among their amounts of change. Weighting the different series of changing prices by the reciprocals of base-year magnitudes overweights rising prices as compared with falling prices, prices that rise rapidly as compared with prices that rise slowly, and prices that fall slowly as compared with prices that fall rapidly. With the harmonic average, which utilizes the reciprocals of ordinary price relatives, the bias of the weighting is inverted.

These elementary principles account both for substantially all the systematic skewness of the distribution of price relatives,[1] and for the bias which Professor Fisher attributes to the arithmetic and harmonic averages. Further, that it is the use of a particular form of price relatives and not the use of "ratios" that is at fault, is proved by the fact that it is possible to construct unweighted arithmetic averages of price relatives which have no systematic bias.

For this purpose the mean proportionals between prices in the two years to be compared must be taken as bases. The changes of different prices are in this manner made accurately comparable. The index number (arithmetic) for the second year as compared with the first is

$$\frac{\Sigma \frac{p_2}{\sqrt{p_1 p_2}}}{n} \div \frac{\Sigma \frac{p_1}{\sqrt{p_1 p_2}}}{n}, \text{ or } \Sigma \sqrt{\frac{p_2}{p_1}} \div \Sigma \sqrt{\frac{p_1}{p_2}}$$

This index number has a number of interesting properties. It meets the time-reversal (but not the circular) test. It has no systematic bias. Harmonic averages give precisely the same results as arithmetic. In general, it will agree very closely with the geometric average. If only two series of prices are used, it is identical with the geometric mean. It is probably the only defensible unweighted index number constructed

[1] Professor Fisher ascribes the skewness of the dispersion of price relatives to the circumstance that "there is more room for dispersion upward." "In the downward direction they are limited by zero, while upward there is no limit." That explanation hardly goes to the roots of the matter. The *range* of dispersion is one thing. The *distribution* of dispersion within its range is quite another thing. In constructing his adjusted index of physical production, Professor E. E. Day found that the skewness of the distribution of his relatives, using the "normal" as a base, was *downward*. Yet there was "more room" upward.

by using arithmetic or harmonic averages of "relatives." In fact, for year-with-year comparisons, it seems to be the best of the unweighted index numbers.[1] For the construction of series of unweighted index numbers the geometric average, conforming as it does to the circular test, retains its superiority.

Professor Fisher throughout seems to give rather grudging recognition to the good qualities of the geometric average. For example, he ranks the unweighted geometric average below the median. The reason he assigns for this rather surprising judgment is that the median, by his tests, agrees just a little more closely with the "ideal" aggregative formula than the geometric does. He observes: "When we assume that simple or equal weighting is the right weighting, the order of merit would make the geometric best and the median far inferior. But, of course, simple weighting never really is the right weighting, and our table of merit is based not on simple but on true weighting." I do not find this reasoning convincing. *One* test is not sufficient. The degree of direct or inverse correlation of weights and prices is subject to variation, as are the various accidental circumstances that make weighted and unweighted index numbers differ. To prove his point Professor Fisher would have to show that *in general* the unweighted median will agree more closely with an accurate weighted index number than will the unweighted geometric. Lacking such proof, the various unweighted index numbers must be compared as among themselves, and on that basis, as Professor Fisher says, the geometric appears to be distinctly superior to the median.

[1] Properly weighted, it takes the aggregative form, $\frac{\sqrt{q_0 q_1}\, p_1}{\sqrt{q_0 q_1}\, p_0}$

Consistently with the small importance he attaches to skewness, Professor Fisher gives no weight to the argument for the geometric average which several writers have built on the circumstance that the logarithms of price relatives are quite generally distributed with a fair approach to symmetry around their arithmetic average, that is, around the geometric average of the price relatives themselves. This argument has some significance, but it is easy to overstress it. The geometric distribution of price relatives is really arbitrarily created by the way in which those relatives are computed. Its presence indicates merely a random distribution of rates of change.[1] It betters the case for the geometric mean only in a negative sort of way. It rules out the common way of using the arithmetic average, which operates unimaginatively upon price relatives as if they were solid quantities. The geometric average sees through their pretense and operates upon the real ratios lying underneath. It should be remembered that the geometric average is wholly independent of the base upon which the relatives are computed, while the character of the distribution depends upon the base. In short, while the arithmetic (and harmonic) averages are distorted by the purely artificial skewness of the distribution, the geometric average is not affected by it. Nor, for that matter, is the median. The character of the distribution of the true rates at which prices change in successive years has a more important bearing upon the appropriateness of the geometric average than the distribution of relative prices has.

The most important arguments for the unweighted

[1] It is easy to prove (although, so far as I know, the point has escaped the attention of students of index numbers) that a normal distribution of rates of change (and of changes of secondary and higher orders) will give a normal distribution of the *logarithms* of price relatives.

geometric average are based upon its general symmetry [1] and upon the weaknesses of its competitors. As compared with the arithmetic mean on the mean-proportional base, its only important advantage seems to be its ability to meet the circular test. That these two averages give results that agree closely constitutes an argument for each of them.[2]

I have been forced to discard the view that there is a substantial difference between measuring an "average change of prices" and measuring a change of the "general level of prices." [3] In a way Professor Fisher is right in holding that "all true index numbers are averages of ratios." But I should prefer to say that all true index numbers are at once averages of ratios *and* ratios of aggregates. This point has a bearing upon a virtue sometimes imputed to the geometric average.

That average, it is held, is in a peculiar way an appropriate average of rates of change. Now the geometric is undoubtedly a true average of ratios, that is, of rates of change, but it has no exclusive prerogative. An arithmetic or harmonic average, as we have seen, may likewise give a true result, and, except for its erratic qualities, so does the median. Care in weighting and in the details of construction are more important than

[1] Including the points that it conforms to the circular test and that the data upon which it operates may be combined first horizontally and then vertically, or in reversed order.

[2] The most serious doubt respecting the geometric average relates to the way in which "geometric compensation" (a matter into which Mr. C. M. Walsh has probed deeply) operates when dispersion is seriously distorted. On general grounds arithmetic averaging seems to be indicated in such cases. But Professor Fisher's severe tests show only unimportant differences in the behavior of the geometric average and of the arithmetic average on the mean-proportional base. [See *Quarterly Journal of Economics*, vol. xxxvii, p. 749.]

[3] This involves a modification of the program, not of the conclusions, of my paper on "The Measurement of Changes of the General Price Level," [*supra*, p. 261].

the choice of one particular type of average rather than another. When properly constructed and properly weighted, averages of ratios, ratios of averages,[1] and ratios of aggregates all come to be about the same thing. Arithmetic, harmonic, geometric, and aggregative types, weighted and used with due regard to their structural peculiarities, agree extraordinarily well in the results they give. That the harmonic average is always smaller than the geometric, and the geometric smaller than the arithmetic, does not, in general, affect the choice of an average for use in making *comparisons*, that is, in constructing index numbers.

The notion that the geometric average is the one proper instrument for measuring rates of change is based, I think, on two misconceptions. In the first place, it has been assumed that an index number of "rates of change" must be an average of fixed-base price relatives — an assumption which at once gives an advantage to the geometric average. In the second place, it has been assumed, mistakenly again, that an average of rates of change must be an *unweighted* index number.

In this latter view the problem of index numbers was held merely to be a special case under the theory of errors. The movement of the "value of money" was to be ascertained by determining the most probable

[1 Professor Fisher replies (*Quarterly Journal of Economics*, vol. xxxvii, p. 743): "A ratio of two averages, each representing the 'price level' of one of the two years compared is an arbitrary shifting figure, an absurdity. Such a ratio is indeterminate." Professor Fisher appears to have lost sight of two considerations: (1) *P*, in his own formulation of the equation of exchange, is an average of actual prices, not an average of what he calls "price ratios." (2) The "ideal" formula, which he places first among all index-number formulas, can be resolved into *either* an average of ratios or a ratio of averages. One process is neither no more nor no less abitrary than the other. See above, pp. 274 and 286.]

general or net movement of a collection of prices, the movement of each price being viewed as a single independent observation on the movement of the value of money.[1] There are two considerations that weigh against this view. (1) The division of goods into "commodities" is more or less arbitrary. The price of each unit of every good is itself a report of an observation upon the "value of money." (2) The problem of index numbers, as Mr. C. M. Walsh has shown, does not fall in the field of the theory of errors.[2]

The truth seems to be that properly constructed averages of "rates of changes" are identical with properly made comparisons of "general price levels" or "general purchasing power." The old Jevons-Laspeyres controversy [3] appears to be drawing to an end, as closer approximations to the two goals indicate that they are at a common point. It is also being discovered that, from whatever direction the common goal be approached, it is not quite attainable. A theoretically perfect index number is an impossibility.

As a practical problem, intimately related to practical questions of policy and control, the problem of index numbers is capable of a practical working solution. As a problem in pure theory, it cannot be completely solved. The difficulty is that "the average change of prices," "changes in the purchasing power of money,"

[1] Such, for example, was the view Cournot and Jevons took.

[2] Excepting, of course, questions relating to the adequacy of the sample.

[3] "It has not been clear whether index numbers really represent measures of a composite quantity, or whether they are probable estimates of the value of a single quantity formed by combining a number of independent observations towards the value of this quantity. The original Jevonian conception of an index number of the value of money was decidedly of the latter type. Modern work on the subject has been increasingly dominated by the other conception." — J. M. Keynes, *A Treatise on Probability*, p. 212.

etc., are practical, not empirical, categories. Concrete facts do not quite fit into them. The composition of the output of goods for which money is exchanged is continually changing. The price level of to-day and the price level of yesterday are incommensurables. Changes of the value of money are qualitative as well as quantitative, and the qualitative changes elude our measurements. It happens, however, that qualitative changes come more slowly than the quantitative ones, so that index numbers may be made quite as precise as their practical uses can possibly demand.

Professor Fisher uses his time-reversal and factor-reversal tests as instruments for finding new formulas. When a formula does not meet the time-reversal test, it must have a "time-antithesis" with an equal but opposed bias, which will be the reciprocal of the formula with its base reversed. Thus the unweighted harmonic average of price relatives is the time antithesis of the arithmetic. The time antitheses of the geometric and of the aggregative weighted by basing-year "values" are the identical formulas weighted by stated-year values. The "factor antithesis" of a given formula is obtained by interchanging the p's and q's and then dividing it into the "value ratio," $\Sigma q_1 p_1 / \Sigma q_0 p_0$. Most of the factor antitheses of standard formulas are themselves completely new formulas. The most notable exception is $\Sigma q_1 p_1 / \Sigma q_1 \div \Sigma q_0 p_0 / \Sigma q_0$, which has been used in making index numbers of the prices of exports and imports. The two fundamental aggregative formulas (Laspeyres's and Paasche's) are the factor antitheses, as they are the time antitheses, of each other, and are the only formulas in which these two properties are joined.

Professor Fisher's next step is to "rectify" all of his biased formulas by "crossing" them with their respective time or factor antitheses. The rectified formulas are the geometric means of the antithetical formulas. The time-crossed formulas all meet the time-reversal test; the factor-crossed formulas all obey the factor-reversal test. Only the "ideal" formula, compounded of the two fundamental aggregative formulas, is on both lists. So the final step is "double rectification" by time crosses of factor-crossed formulas or factor crosses of time-crossed formulas. A doubly rectified formula is thus the fourth root of the product of a fundamental formula, its time antithesis, and the factor antitheses of the two.

The book is built around the time-reversal and factor-reversal tests and the developments of them which have just been described. The whole scheme constitutes an important and unique addition to the stock of knowledge of index numbers. This new method of attack, together with his careful experimental tests, are Professor Fisher's largest contributions. Their greatest value, I think, is that they clear the field of a number of inaccurate forms and center attention upon a few trustworthy forms and a few important problems.

In view of the great importance of these contributions, it may be ungracious to suggest that Professor Fisher has wasted a good deal of labor. With uncompromising thoroughness he has put all the twenty-six formulas with which he starts through the different steps of finding time and factor antitheses, rectification, double rectification, and testing. He even goes so far as to find and rectify the factor antitheses of unweighted formulas. Many of the more accurate formulas that he develops are unwieldy and impracticable. None of the rectified

formulas meet the circular test. A few of them, as Professor Fisher shows, are not true averages. Moreover, although it leads undeniably to important results, rectification by crossing is not the only possible way of developing good index numbers.

By Professor Fisher's method, when biased error is found a compensating error is created and the two errors are combined in such a way that they cancel each other. Thus the doubly-rectified formulas contain four separate elements of biased error, arranged in compensating pairs. Why introduce avoidable error? Instead of finding its double, why not eliminate it? It is true, of course, that to cancel error is to eliminate it, but there is no certainty that the process will point to the nature and the sources of error, or contribute to the improvement of the general theory of index numbers. In the special case of the aggregative type, however, rectification unmistakably justifies itself by its fruits. For it points unequivocally to the "ideal" form — the geometric mean of the two fundamental aggregative types, which is, beyond much doubt, the most accurate single index number of the movement of prices between any two years. Its closest rival among the rectified index numbers is the doubly-rectified weighted geometric.

The use of crossed weights, that is, of weights which are means of base-year weights and stated-year weights, is a more familiar process. Professor Fisher includes a study of the possibilities of this method, but, as he explains, only "in deference to the wishes of other students of index numbers, and in order that the list shall cover all formulæ previously suggested by others and all points of view." It is fortunate that Professor Fisher has allowed himself to be persuaded to take account of these cross-weighted formulas. For he finds that

"crossing the weights of two formulæ of the same model, and so forming a new formula, yields almost identically the same numerical results as crossing the formulæ themselves." But, "formulæ crossing is a universal method of compromising between two formulæ, while weight crossing is of restricted importance."

Now an examination of Professor Fisher's results suggests that the difference between crossing formulas and crossing weights may be expressed in another way. Where the original formula contains bias due to other sources than biased weighting, formula crossing will, while weight crossing will not, eliminate the bias. Thus weight crossing cannot eliminate the bias put into arithmetic and harmonic means by the improper use of price relatives. But when the original formulas are inherently sound, weight crossing gives substantially as good results as formula crossing. Thus the cross-weighted geometrics and aggregatives are excellent index numbers.

The most interesting of the cross-weighted formulas is $\frac{\Sigma\sqrt{q_0q_1}p_1}{\Sigma\sqrt{q_0q_1}p_0}$. Professor Fisher shows, in an instructive bit of analysis,[1] that it is probably not quite so reliable as the simpler form, $\frac{\Sigma(q_0+q_1)p_1}{\Sigma(q_0+q_1)p_0}$, a formula which has the approval of Professors Marshall and Edgeworth and Mr. Walsh. This formula and the "ideal" formula agree very closely in their results. Because it is simple and easily calculated, Professor

[1] But his proof that it is better to take the arithmetic than the geometric mean of the weights holds only for aggregative, not for geometric, index numbers.

Fisher recommends it, instead of the "ideal," for general use. I, for one, accept this conclusion, but with two reservations. First, where consistent series rather than the most accurate year-with-year comparisons are desired, the use of the simple aggregative $\Sigma q_0 p_1 / \Sigma q_0 p_0$, with weights revised from time to time, is preferable. Second, there is little difference between the accuracy of aggregative and geometric types. Peculiarities of the available data will often indicate that, in some particular use, the geometric should be preferred.

530.1
Y84

RAYMOND H. FOGLER LIBRARY
DATE DUE

OCT 23 1972	

UNIV OF MAINE
3 7550 00895119 7